arXiv: arXiv:0000.0000

ADAPTIVE ESTIMATION UNDER SINGLE-INDEX CONSTRAINT IN A REGRESSION MODEL *

By Oleg Lepski and Nora Serdyukova

Aix-Marseille Université and Georg-August-Universität Göttingen

The problem of adaptive multivariate function estimation under single-index assumption is studied in the framework of the regression model with random design. We consider the case when both the link function and index vector are unknown. We propose a novel estimation procedure that adapts simultaneously to the unknown index vector and the smoothness of link function by selecting from a family of specific kernel estimators. We establish a pointwise oracle inequality which, in its turn, is used to judge the quality of estimating the entire function (global oracle inequality). Both results are applied to the problems of pointwise and global adaptive estimation over a collection of Hölder and Nikol'skii functional classes.

1. Introduction.

This paper deals with multivariate functions estimation. We establish local as well as global oracle inequalities and show how to use them for deriving minimax adaptive results.

Model and set-up. We observe data $(X_1, Y_1), \dots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$,

$$Y_i = F(X_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $d \geq 2$, the noise $\{\varepsilon_i\}_{i=1}^n$ are i.i.d centered random variables, satisfying moment conditions given in Assumption 1 below, and the design points $\{X_i\}_{i=1}^n$ are independent random vectors with common density g with respect to the Lebesgue measure. The sequences $\{\varepsilon_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ are assumed to be independent, and the density g is supposed to be known.

Additionally, we assume that the function $F : \mathbb{R}^d \to \mathbb{R}$ possesses a single-index structure, that is there exist $f : \mathbb{R} \to \mathbb{R}$ and $\theta^* \in \mathbb{R}^d$ such that

$$F(x) = f(x^\top \theta^*). \tag{1.2}$$

A minimal technical assumption imposed on the link function is that f belongs to some Hölder ball. We would like to emphasize that the knowledge

*Funding of the ANR-07-BLAN-0234 is acknowledged. The second author is also supported by the DFG FOR 916.

AMS 2000 subject classifications: Primary 62G05; secondary 62G08, 62G20

Keywords and phrases: Adaptive estimation, lower bounds, minimax rate of convergence, nonparametric function estimation, nonparametric regression, oracle inequalities, single-index model, structural adaptation.

of this ball will not be required, in particular this information is not used in the proposed estimation procedure (for more detail, see the discussion after Assumption 3.) All the results established in the paper, except the lower bound proved in Theorem 4, are obtained for $d = 2$. The principal difficulties in extending these results to the case of an arbitrary dimension are discussed in Remark 2. We note also, that the single-index assumption even if $d = 2$ is a direct generalization of the univariate regression model. As a consequence, our results, mainly presented in Section 2.2, generalize in several directions existing ones obtained in the univariate regression model with random design, see discussion after Theorem 5 and the references therein.

Thus, the paper aims at estimating the entire function F on $[-1/2, 1/2]^2$ or its value $F(x)$ from the data $\{(X_i, Y_i)\}_{i=1}^n$ without any prior knowledge of the nuisance parameters $f(\cdot)$ and θ^*. The unit square is chosen for the notation convenience, and all the results remain true when $[-1/2, 1/2]^2$ is replaced by an arbitrary bounded interval of $\mathbb{R}^2$.

Throughout the paper we adopt the following notations. The joint distribution of the sequence $\{(X_i, Y_i)\}_{i=1}^n$ will be denoted by $\mathbb{P}_F^{(n)}$, those of $\{(X_i, \varepsilon_i)\}_{i=1}^n$ by $\mathbb{P}_{X,\varepsilon}^{(n)}$. Moreover $\mathbb{P}_X^{(n)}$ and $\mathbb{P}_\varepsilon^{(n)}$ will be used for marginal distribution of $\{X_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ respectively.

To judge the quality of estimation we use either the risk determined by the L_r norm, $\|\cdot\|_r$, on $[-1/2, 1/2]^2$ with $r \in [1, \infty)$ ("global" risk)

$$\mathcal{R}_r^{(n)}(\widehat{F}, F) = \mathbb{E}_F^{(n)} \|\widehat{F}(\cdot) - F(\cdot)\|_r, \tag{1.3}$$

or the "pointwise" risk

$$\mathcal{R}_{r,x}^{(n)}(\widehat{F}, F) = (\mathbb{E}_F^{(n)} |\widehat{F}(x) - F(x)|^r)^{1/r}, \quad x \in [-1/2, 1/2]^2. \tag{1.4}$$

Here $\widehat{F}(\cdot)$ is an estimator, i.e. an $\{(X_i, Y_i)\}_{i=1}^n$-measurable function and $\mathbb{E}_F^{(n)}$ denotes the mathematical expectation with respect to $\mathbb{P}_F^{(n)}$.

Main assumptions. Let us formulate the principal assumptions used in the sequel. They are imposed on the distributions of the design and noise variables as well as on the approximation property of the link function

Assumption 1. *The random variable ε_1 has a symmetric distribution with density p with respect to the Lebesgue measure. Moreover, there exist $\Upsilon > 0$, $\Omega \in (0, 1]$, and $\omega > 0$ such that*

$$p \in \mathfrak{P} := \left\{ \ell : \mathbb{R} \to \mathbb{R}_+ : \int_x^\infty \ell(y) \mathrm{d}y \le \Upsilon e^{-\Omega x^\omega}, \ \forall x \ge 0 \right\}.$$

The assumption holds, for example, for the Gaussian, Laplace or more generally for symmetrized Weibull distribution. Throughout the paper the functional class $\mathfrak{P}$ is supposed to be fixed.

ASSUMPTION 2. *There exists* $\underline{g} \in (0,1]$ *such that* $\inf_{x\in[-5/2,5/2]^2} g(x) \geq \underline{g}$.

The assumption holds obviously if the design points are uniformly distributed on any bounded Borel set containing $[-5/2, 5/2]^2$.

Note that imposed condition is "fitted" to the estimation over $[-1/2, 1/2]^2$ that motivates the appearance of the set $[-5/2, 5/2]^2$. In general case, when the estimation problem is considered on some rectangle $[a, b] \times [c, e] \in \mathbb{R}^2$, the corresponding infimum should be taken over the rectangle $[a - 2, b + 2] \times [c - 2, e + 2]$. We remark that independently of the values a, b, c, e the assumption will be fulfilled if $g \in \mathbb{C}\big(\mathbb{R}^2\big)$ and $g(x) > 0$ for any $x \in \mathbb{R}^2$.

ASSUMPTION 3. *There exist* $\beta_0 \in (0, 1)$ *and* $M > 0$ *such that*

$$f \in \mathbb{F}(\beta_0, M) := \left\{ U : \mathbb{R} \to \mathbb{R} : \quad \|U\|_\infty + \sup_{y_1,y_2\in\mathbb{R}} \frac{|U(y_1) - U(y_2)|}{|y_1 - y_2|^{\beta_0}} \leq M \right\}.$$

The latter assumption guarantees that the link function is smooth. However, it is important to emphasize that the parameters β_0 and M are not supposed to be known *a priory*. In particular, they are not involved in the estimation procedure developed in the paper. On the other hand, both parameters restrict the minimal sample size needed to justify all the theoretical results. Set for any $n \in \mathbb{N}^*$

$$h_{\min} = n^{-1} \ln^{1+\frac{2}{\omega}}(n), \qquad \mathfrak{h} = \sqrt{n^{-1} \ln^{1+\frac{1}{\omega}}(n)}. \tag{1.5}$$

In the sequel it will be assumed that $n \geq n_0$, where

$$n_0 = \inf \left\{ m \in \mathbb{N}^* : \ (M \vee 1) \max\left(\mathfrak{h}^{\beta_0},\ \ln^{\frac{1}{\omega}}(n) h_{\min}^{\beta_0}\right) \leq 1,\ \forall n \geq m \right\}. \tag{1.6}$$

We finish this discussion with the following remark. All the results obtained in the paper remain true if one assumes that $f \in \mathbb{F}(0, M)$ (the uniform boundedness of the link function) and M is known.

Objectives. For clarity of presentation, we will assume that the index vector $\theta^* \in \mathbb{S}^1$, where $\mathbb{S}^{d-1}$ is the unite sphere in $\mathbb{R}^d$. However, in Section 2.1.4 it is shown that our results can be extended to the case $\theta^* \in \mathbb{R}^2$.

The goal of our studies is at least threefold. First, we seek an estimation procedure $\widehat{F}(x), x \in [-1/2, 1/2]^2$, for F which could be applicable to any

function F satisfying (1.2). Moreover, we would like to bound the risk of this estimator uniformly over the set $\mathbb{F}(\beta_0, M)\times\mathbb{S}^1$. More precisely, we want to establish for $\widehat{F}(x)$ the so-called local oracle inequality: for any $x \in [-1/2, 1/2]^2$

$$\mathcal{R}^{(n)}_{r,x}(\widehat{F}, F) \le C_r A^{(n)}_{f,\theta^*}(x), \quad \forall f \in \mathbb{F}(\beta_0, M), \;\; \forall \theta^* \in \mathbb{S}^1. \tag{1.7}$$

Here the quantity $A^{(n)}_{f,\theta^*}(\cdot)$ is completely determined by the function f, vector θ^* and observations number n, while C_r is a numerical constant independent of F and n.

Being established the local oracle inequality allows us to derive minimax adaptive results for the function estimation at a given point. Indeed, let $\mathbb{F}(\gamma), \;\gamma \in \Gamma$, be a collection of functional classes such that $\cup_{\gamma\in\Gamma}\mathbb{F}(\gamma) \subseteq \mathbb{F}(\beta_0, M)$. For any $\gamma \in \Gamma$ define

$$\phi_n(\gamma) = \inf_{\widetilde{F}} \sup_{(f,\theta^*)\in\mathbb{F}(\gamma)\times\mathbb{S}^1} \mathcal{R}^{(n)}_{r,x}(\widetilde{F}, F),$$

where infimum is taken over all possible estimators. The quantity $\phi_n(\gamma)$ is the minimax risk on $\mathbb{F}(\gamma)\times\mathbb{S}^1$, and the problem arisen in the framework of minimax adaptive estimation consists in the following. One has to construct an estimator F^* such that for any $\gamma \in \Gamma$

$$\sup_{(f,\theta^*)\in\mathbb{F}(\gamma)\times\mathbb{S}^1} \mathcal{R}^{(n)}_{r,x}(F^*, F) \;\asymp\; \phi_n(\gamma), \;\; n \to \infty. \tag{1.8}$$

The estimator F^* satisfying (1.8) is called optimally rate adaptive over the collection $\{\mathbb{F}(\gamma), \;\gamma \in \Gamma\}$. Let (1.7) be proved and suppose that for any $\gamma \in \Gamma$

$$\sup_{(f,\theta^*)\in\mathbb{F}(\gamma)\times\mathbb{S}^1} A^{(n)}_{f,\theta^*}(x) \asymp\; \phi_n(\gamma), \;\; n \to \infty.$$

Then one can assert that the estimator $\widehat{F}$ is adaptive over $\{\mathbb{F}(\gamma), \;\gamma \in \Gamma\}$.

Thus, the first task is to prove (1.7). To the best of our knowledge such kind results do not exist in the context of the regression with random design not only under the single-index constraint, but also in a univariate regression.

Next, we apply this result to minimax adaptive estimation over the collection $\mathbb{F}(\gamma) = \mathbb{H}(\beta, L), \;\gamma = (\beta, L)$, where $\mathbb{H}(\beta, L)$ is a Hölder class of functions, see Definition 1. In particular, we find the minimax rate over $\mathbb{H}(\beta, L)\times\mathbb{S}^1$ and prove that our estimator $\widehat{F}$ achieves that rate, i.e. is optimally adaptive. This result is quite surprising because, if θ^* is fixed, say $\theta^* = (1, 0)^\top$, then it is well known that an optimally adaptive estimator does not exist, see Lepski (1990) (Gaussian white noise model), Brown and Low (1996) (density model) and Gaïffas (2007) (regression model).

Note also that local oracle inequality (1.7) allows us to bound from above the "global" risk as well. Indeed, in view of Jensen's inequality and Fubini's theorem

$$\Big[\mathcal{R}_r^{(n)}(\widehat{F},F)\Big]^r \leq \mathbb{E}_F^{(n)}\big\|\widehat{F}(\cdot)-F(\cdot)\big\|_r^r = \big\|\mathcal{R}_{r,\cdot}^{(n)}(\widehat{F},F)\big\|_r^r$$

and, therefore,

$$\mathcal{R}_r^{(n)}(\widehat{F},F) \leq C_r\big\|A_{f,\theta^*}^{(n)}\big\|_r. \tag{1.9}$$

The latter inequality is called global oracle inequality, and in the framework of the present study it supplies new results. As local oracle inequality (1.7) is a powerful tool for deriving minimax adaptive results in pointwise estimation so global oracle inequality (1.9) can be used for constructing adaptive estimators of the entire function F.

We will consider the collection of Nikol'skii classes $\mathbb{N}_p(\beta,L)$, see Definition 2, where $\beta, L>0$ and $1\leq p<\infty$. It is important to emphasize that by considering these classes we want to estimate the functions with inhomogeneous smoothness. This means that the underlying function can be very regular on some parts of the observation domain and rather irregular on the other sets.

We will compute the asymptotic bounds on

$$\sup_{(f,\theta^*)\in\mathbb{N}_p(\beta,L)\times\mathbb{S}^1}\big\|A_{f,\theta^*}^{(n)}\big\|_r$$

and show that, if $(2\beta+1)p<r$, the rate of convergence coincides with the minimax rate over $\mathbb{N}_p(\beta,L)\times\mathbb{S}^1$. This means that our estimator $\widehat{F}$ is optimally rate adaptive over the collection $\big\{\mathbb{N}_p(\beta,L)\times\mathbb{S}^1,\ \beta>0, L>0\big\}$ whenever $(2\beta+1)p<r$. In the case $(2\beta+1)p\geq r$ we will prove that the latter bound differs from the bound on the minimax risk by a logarithmic factor. Following the contemporary language we say that the estimator $\widehat{F}$ is "nearly" adaptive. However, the construction of an optimally rate adaptive over the entire range of Nikol'skii classes estimator under single-index constraint (1.2) remains an open problem.

We would like to emphasize that all these results are completely new. The adaptive estimation under the L_r loss and single-index constraint, except the case $r=2$, Gaïffas and Lecué (2007), was not studied. Note, however, that the cited results were obtained under the Gaussian errors model and over the collection of Hölder classes that does not admit the consideration of inhomogeneous functions.

Remarks. It turns out that the adaptation to the unknown θ^* and $f(\cdot)$ can be formulated in terms of selection from a special family of kernel estimators in the spirit of the Lepski and the Goldenschluger-Lepski selection rules, see Lepski (1990), Kerkyacharian et al. (2001), Goldenshluger and Lepski (2008). However, the proposed here procedure is quite different from the aforementioned ones, and it allows us to solve the problem of minimax adaptive estimation under the L_r losses over a collection of Nikol'skii classes.

It is worth mentioning that the considered single-index model is not only of high theoretical interest but is also actively used especially in econometrics, e.g. Horowitz (1998), Maddala (1983). The estimation, nevertheless, is usually performed under smoothness assumptions on the link function. One usually uses the L_2 losses, and the available methodology is based on these restrictions. To the best of our knowledge the only exceptions are Golubev (1992) for the minimax estimation under the projection pursuit constraints, and Goldenshluger and Lepski (2009) presenting a novel procedure permitting to adapt simultaneously to unknown smoothness and structure.

Organization of the paper. In Section 2 we motivate and explain the proposed selection rule and establish for it local and global oracle inequalities, Section 2.1. Section 2.2 is devoted to the application of these results to minimax adaptive estimation. The proofs of the main results are given in Section 3, and the proofs of technical lemmas are postponed until Appendix.

2. Main results.

In this section we present our procedure and establish for it local and global oracle inequalities. Then, we apply these results to adaptive estimation over a collection of Hölder classes (pointwise estimation) and over a collection of Nikol'skii classes (estimating the entire function with the accuracy of an estimator measured under the L_r risk).

2.1. *Oracle approach.*

Let $\mathcal{K}:\mathbb{R}\to\mathbb{R}$ be a function (kernel) satisfying condition $\int\mathcal{K}=1$. With any such $\mathcal{K}$, any $z\in\mathbb{R}$, $h\in(0,1]$ and any $f\in\mathbb{F}(\beta_0,M)$ we associate the quantity

$$\Delta_{\mathcal{K},f}(h,z)=\sup_{\delta\le h}\left|\delta^{-1}\int\mathcal{K}\big([u-z]/\delta\big)\big(f(u)-f(z)\big)\mathrm{d}u\right|.$$

We note that $1/\delta\int\mathcal{K}\big([u-z]/\delta\big)f(u)\mathrm{d}u$ (kernel smoother) can be understood as an approximation of the function f at the point z. Thus, $\Delta_{\mathcal{K},f}(h,z)$ is a monotonous approximation error provided by this kernel smoother. In particular, $\Delta_{\mathcal{K},f}(h,z)\to 0$ as $h\to 0$ in view of Assumption 3.

Throughout the paper $\|\mathcal{K}\|_p$, $1\le p\le\infty$, denotes the L_p norm of $\mathcal{K}$ and we will assume that the kernel $\mathcal{K}$ satisfies the following condition.

Assumption 4. *(1) supp*$(\mathcal{K}) \subseteq [-1/2, 1/2]$, $\int \mathcal{K} = 1$, $\mathcal{K}$ *is symmetric;*

(2) there exists $Q > 0$ *such that*

$$\big|\mathcal{K}(u) - \mathcal{K}(v)\big| \le Q|u - v|, \ \forall u, v \in \mathbb{R}\,.$$

2.1.1. *Oracle estimator.* For any $y \in \mathbb{R}$ define

$$\Delta^*_{\mathcal{K},f}(h, y) = \sup_{a>0}(2a)^{-1}\int_{y-a}^{y+a}\Delta_{\mathcal{K},f}(h, z)\mathrm{d}z.$$

Thus, $\Delta^*_{\mathcal{K},f}(h, \cdot)$ is the Hardy-Littlewood maximal function of $\Delta_{\mathcal{K},f}(h, \cdot)$, see for example Wheeden and Zygmund (1977). Note that since $f \in \mathbb{F}(\beta_0, M)$ so $\Delta^*_{\mathcal{K},f}(h, \cdot) \ge \Delta_{\mathcal{K},f}(h, \cdot)$.
Now we are in a position to introduce the oracle estimator. Set for any $y \in \mathbb{R}$

$$h^*_{\mathcal{K},f}(y) = \sup\Big\{h \in \big[h_{\min}, 1\big] : \sqrt{nh}\,\Delta^*_{\mathcal{K},f}(h, y) \le \|\mathcal{K}\|_\infty\sqrt{\ln(n)}\Big\}, \tag{2.1}$$

where $h_{\min}$ is defined in (1.5).

Some remarks are in order. First, we note that $\Delta^*_{\mathcal{K},f}(h, \cdot) \le M\|\mathcal{K}\|_1 h^{\beta_0}$ for any $f \in \mathbb{F}(\beta_0, M)$ and any $h > 0$. Hence, $\sqrt{nh_{\min}}\Delta^*_{\mathcal{K},f}(h_{\min}, \cdot) \le \|\mathcal{K}\|_1\sqrt{\ln(n)}$ for any $n \ge n_0$ in view of (1.6). Next, Assumption 4 (*2*) implies obviously that $\Delta^*_{\mathcal{K},f}(\cdot, y)$ is a continuous function and, hence,

$$\text{either} \qquad \sqrt{nh^*_{\mathcal{K},f}(y)}\,\Delta^*_{\mathcal{K},f}\big(h^*_{\mathcal{K},f}(y), y\big) = \|\mathcal{K}\|_\infty\sqrt{\ln(n)}, \tag{2.2}$$

$$\text{or} \qquad \sqrt{nh}\,\Delta^*_{\mathcal{K},f}(h, y) \le \|\mathcal{K}\|_\infty\sqrt{\ln(n)}, \ \ \forall h \in [h_{\min}, 1]. \tag{2.3}$$

Here we have also used that $\|\mathcal{K}\|_1 \le \|\mathcal{K}\|_\infty$ in view of Assumption 4 (*1*).

The quantity similar to the defined above $h^*_{\mathcal{K},f}(\cdot)$ first appeared in Lepski et al. (1997) in the context of the estimating univariate functions possessing inhomogeneous smoothness. Some years later this approach has been developed in Kerkyacharian et al. (2001) and Goldenshluger and Lepski (2008) for the multivariate function estimation. In these papers, the interested reader can find a more detailed discussion of the oracle approach. In the present paper we try to adopt the "ideology" proposed in the aforementioned papers to the estimation under single-index constraint. Our main idea is based on the following rather simple observation:

For any $(\theta, h) \in \mathbb{S}^1 \times [h_{\min}, 1]$ define the matrix

$$E_{(\theta,h)} = \begin{pmatrix} h^{-1}\theta_1 & h^{-1}\theta_2 \\ -\theta_2 & \theta_1 \end{pmatrix}$$

and consider the family of kernel estimators

$$\mathcal{F}=\Big\{\widehat{F}_{(\theta,h)}(\cdot)=\frac{\det\big(E_{(\theta,h)}\big)}{n}\sum_{i=1}^{n}\frac{K\big(E_{(\theta,h)}(X_i-\cdot)\big)}{g(X_i)}Y_i,\ (\theta,h)\in\mathbb{S}^1\times[h_{\min},1]\Big\}.$$

Here $K(u,v)=\mathcal{K}(u)\mathcal{K}(v)$. We remark, first, that in view of Assumption 2 the estimator $\widehat{F}_{(\theta,h)}$ is well defined since

$$K\big(E_{(\theta,h)}(t-x)\big)=0,\quad \forall t\in[-3/2,3/2],\ \ \forall x\in[-1/2,1/2].$$

This property follows from Assumption 4 (*1*). Moreover, $\det\big(E_{(\theta,h)}\big)=h^{-1}$.

The choice $\theta=\theta^*$ and $h=h^*:=h^*_{\mathcal{K},f}(x^T\theta^*)$ leads to the so-called *oracle estimator* $\widehat{F}_{(\theta^*,h^*)}(\cdot)$. First, we note that $\widehat{F}_{(\theta^*,h^*)}(\cdot)$ is not an estimator in the usual sense, since it depends on the function F to be estimated (more precisely on (f,θ^*) which determines F). The meaning of this estimator is explained by the following result.

PROPOSITION 1. *For any $(f,\theta^*)\in\mathbb{F}(\beta_0,M)\times\mathbb{S}^1$, $r\geq 1$ and any $n\geq n_0$*

$$\mathcal{R}^{(n)}_{r,x}\big(\widehat{F}_{(\theta^*,h^*)},F\big)\leq c_r\left[\frac{\ln(n)}{nh^*_{\mathcal{K},f}(x^\top\theta^*)}\right]^{\frac{1}{2}},\forall x\in[-1/2,1/2]^2,$$

where $c_r>0$ is a numerical constant independent of n.

The proof of the proposition, based on the Rozenthal inequality, is straightforward and can be omitted.

The result of Proposition 1 can be treated as follows. The “oracle” knows the exact value of the index vector θ^* and the optimal, up to $\ln(n)$, trade-off h^* between the approximation error determined by $\Delta^*_{\mathcal{K},f}(h^*,\cdot)$ and the stochastic error provided by the kernel estimator from the collection $\mathcal{F}$ with bandwidth h^*. It explains why the “oracle” chooses the “estimator” $\widehat{F}_{(\theta^*,h^*)}$.

In the next paragraph we propose a “real” (based on the observation) estimator $\widehat{F}(\cdot)$, which mimics the oracle estimator. This means that for any $(f,\theta^*)\in\mathbb{F}(\beta_0,M)\times\mathbb{S}^1$, $x\in[-1/2,1/2]^2$, $r\geq 1$ and $n\geq n_0$

$$\mathcal{R}^{(n)}_{r,x}\big(\widehat{F},F\big)\leq c'_r\left[\frac{\ln(n)}{nh^*_{\mathcal{K},f}(x^\top\theta^*)}\right]^{\frac{1}{2}},\forall x\in[-1/2,1/2]^2,$$

where c'_r is an absolute constant independent of the number of observations n and the underlying function F. The latter result is a local oracle inequality. The construction of the estimator $\widehat{F}(\cdot)$ is based on the data-driven selection from the family $\mathcal{F}$.

2.1.2. *Selection rule.* For any $\theta, \nu \in \mathbb{S}^1$ and any $h \in [h_{\min}, 1]$ define

$$\overline{E}_{(\theta,h)(\nu,h)} = \begin{pmatrix} \frac{(\theta_1+\nu_1)}{2h(1+|\nu^\top\theta|)} & \frac{(\theta_2+\nu_2)}{2h(1+|\nu^\top\theta|)} \\ -\frac{(\theta_2+\nu_2)}{2(1+|\nu^\top\theta|)} & \frac{(\theta_1+\nu_1)}{2(1+|\nu^\top\theta|)} \end{pmatrix},$$

where

$$E_{(\theta,h)(\nu,h)} = \begin{cases} \overline{E}_{(\theta,h)(\nu,h)}, & \nu^\top\theta \geq 0; \\ \overline{E}_{(-\theta,h)(\nu,h)}, & \nu^\top\theta < 0. \end{cases}$$

It is easy to check that $(4h)^{-1} \leq \det\big(E_{(\theta,h)(\nu,h)}\big) \leq (2h)^{-1}$. A kernel estimator associated with the matrix $E_{(\theta,h)(\nu,h)}$ is defined by

$$\widehat{F}_{(\theta,h)(\nu,h)}(x) = \frac{\det\big(E_{(\theta,h)(\nu,h)}\big)}{n} \sum_{i=1}^{n} \frac{K\big(E_{(\theta,h)(\nu,h)}(X_i - \cdot)\big)}{g(X_i)} Y_i. \tag{2.4}$$

Once again we note that the estimator $\widehat{F}_{(\theta,h)(\nu,h)}$ is well-defined since

$$K\big(E_{(\theta,h)}(t-x)\big) = 0, \quad \forall t \in [-5/2, 5/2], \;\; \forall x \in [-1/2, 1/2].$$

For any $u_1, u_2 \in \mathbb{R}$ set $K_{\mathfrak{h}}(u_1, u_2) = \mathfrak{h}^{-2}\mathcal{K}(u_1/\mathfrak{h})\mathcal{K}(u_2/\mathfrak{h})$ and introduce

$$\widehat{F}(t) = n^{-1}\sum_{i=1}^{n} g^{-1}(X_i)K_{\mathfrak{h}}(X_i - t)Y_i, \qquad \widehat{F}_\infty = 2\|\widehat{F}\|_\infty + 2C_5(n),$$

where $\|\widehat{F}\|_\infty = \sup_{t\in[-1/2,1/2]^2}\big|\widehat{F}(t)\big|$, and $\mathfrak{h}$ is defined in (1.5). Put also

$$\mathrm{TH}(\eta) = 2\Big[\|\mathcal{K}\|_\infty^2\sqrt{\ln(n)} + \widehat{F}_\infty C_1(n) + C_2(n)\Big](\eta n)^{-1/2}, \;\; \eta \in (0,1].$$

The quantities $C_1(n)$, $C_2(n)$ and $C_5(n)$ are listed at the beginning of Section 3.1. Those explicit expressions are too cumbersome and it is not convenient for us to present them right now.

Set $\mathcal{H}_n = \big\{h_k = 2^{-k},\; k = 0, 1, \dots\big\} \cap \big[2^{-1}h_{\min}, 1\big]$ and let for any $\theta \in \mathbb{S}^1$ and $h \in \mathcal{H}_n$

$$\begin{aligned} R_x^{(1)}(\theta,h) &= \sup_{\eta\in\mathcal{H}_n:\,\eta\leq h}\left[\sup_{\nu\in\mathbb{S}^1}\big|\widehat{F}_{(\theta,\eta)(\nu,\eta)}(x) - \widehat{F}_{(\nu,\eta)}(x)\big| - \mathrm{TH}(\eta)\right]_+; \\ R_x^{(2)}(h) &= \sup_{\eta\in\mathcal{H}_n:\,\eta\leq h}\;\sup_{\theta\in\mathbb{S}^1}\Big[\big|\widehat{F}_{(\theta,h)}(x) - \widehat{F}_{(\theta,\eta)}(x)\big| - \mathrm{TH}(\eta)\Big]_+. \end{aligned}$$

Define $(\widehat{\theta}, \widehat{h})$ as a solution of the following minimization problem:

$$\begin{aligned} &R_x^{(1)}(\widehat{\theta},\widehat{h}) + R_x^{(2)}(\widehat{h}) + \mathrm{TH}(\widehat{h}) \\ &= \inf_{(\theta,h)\in\mathbb{S}^1\times\mathcal{H}_n}\Big[R_x^{(1)}(\theta,h) + R_x^{(2)}(h) + \mathrm{TH}(h)\Big]. \end{aligned} \tag{2.5}$$

Our final estimator is $\widehat{F}(x) = \widehat{F}_{(\widehat{\theta},\widehat{h})}(x)$.

REMARK 1. *We note that all random fields involved in the description of selection rule (2.5) are continuous functions on* $\mathbb{S}^1$ *thanks to Assumption 4 (2) and, moreover,* $\mathcal{H}_n$ *is finite. Thus,* $(\widehat{\theta},\widehat{h})$ *is* $\{(X_i,Y_i)\}_{i=1}^n$ *-measurable and* $(\widehat{\theta},\widehat{h})\in\mathbb{S}^1\times\mathcal{H}_n$*, see Jennrich (1969).*

REMARK 2. *Our selection rule (2.5) is defined in the case* $d=2$*. The main difficulty in extending it to* $d>2$ *consists in the construction of the matrix* $E_{(\theta,h)(\nu,h)}$ *for any vectors* $\theta,\nu\in\mathbb{S}^{d-1}$*. Indeed, analyzing the proof of Theorem 1 we remark that the following properties should be fulfilled.*

$$E_{(\theta,h)(\nu,h)}\in\mathcal{E}_{a,A},\quad E_{(\theta,h)(\nu,h)}=\pm E_{(\nu,h)(\theta,h)},\ \ \forall\theta,\nu\in\mathbb{S}^{d-1},\ \ \forall h\in\mathcal{H}_n,$$

where the class of matrices $\mathcal{E}_{a,A}$ *is defined in* (3.2)*. If* $d=2$*, these requirements hold. However, we were not able to construct a class of matrices obeying latter restrictions in the dimension strictly larger than 2. Note nevertheless that if such class would be found our results could be extended to* $d>2$ *without any additional consideration.*

2.1.3. *Local and global oracle inequalities.* To formulate our main results we need to enforce restriction (1.6) imposed on the minimal sample size n. Let $n_1\geq 1$ be defined as follows

$$n_1=\inf\Big\{m\in\mathbb{N}^*:\ \ \big(n\mathfrak{h}^2\big)^{-\frac{1}{2}}C_3(n)\leq 1/2,\ \ \forall n\geq m\Big\}, \tag{2.6}$$

where $\mathfrak{h}$ is defined in (1.5), and $C_3(n)$ is given at the beginning of Section 3.1. All our results below will be proved under the condition $n\geq n_0\vee n_1$.

First, we note that n_1 is well-defined since $\big(n\mathfrak{h}^2\big)^{-1/2}C_3(n)\to 0$ as $n\to\infty$. Next, contrary to restriction (1.6) that relates the sample size n to the quantities β_0, M appeared in Assumption 3, restriction (2.6) links the minimal value of n with the quantity $\underline{g}$ appeared in Assumption 2.

THEOREM 1. *For any* $(f,\theta^*)\in\mathbb{F}(\beta_0)\times\mathbb{S}^1$*,* $x\in[-1/2,1/2]^2$*,* $r>0$ *and* $n\geq n_0\vee n_1$

$$\mathcal{R}^{(n)}_{r,x}\Big(\widehat{F}_{(\widehat{\theta},\widehat{h})},F\Big)\leq c_{r,1}\bigg(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(x^T\theta^*)}\bigg)^{\frac{1}{2}}+c_{r,2}n^{-\frac{1}{2}}.$$

The constants $c_{r,1}$ and $c_{r,2}$ are independent of n and F and their explicit expressions can be extracted from the proof of the theorem.

As already mentioned, the global oracle inequality is obtained by integrating the local oracle inequality. Indeed, using Jensen's inequality and Fubini's

theorem we get $\mathcal{R}_r^{(n)}(\widehat{F},F) \leq \big\|\mathcal{R}_{r,\cdot}^{(n)}(\widehat{F},F)\big\|_r$ so

$$\mathcal{R}_r^{(n)}(\widehat{F},F) \;\leq\; c_{r,1}\bigg\{\int_{[-\frac{1}{2},\frac{1}{2}]^2}\bigg[\frac{\ln(n)}{nh^*_{\mathcal{K},f}(x^T\theta^*)}\bigg]^{\frac{r}{2}}\mathrm{d}x\bigg\}^{\frac{1}{r}} + c_{r,2}n^{-\frac{1}{2}}.$$

Integration by substitution yields

$$\int_{[-\frac{1}{2},\frac{1}{2}]^2}\bigg(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(x^T\theta^*)}\bigg)^{\frac{r}{2}}\mathrm{d}x \leq \int_{-\frac{1}{2}}^{\frac{1}{2}}\bigg(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(z)}\bigg)^{\frac{r}{2}}\mathrm{d}z,$$

that leads to the following bound.

THEOREM 2. *For any $(f,\theta^*)\in\mathbb{F}(\beta_0,M)\times\mathbb{S}^1$, $r\geq 1$ and $n\geq n_0\vee n_1$*

$$\mathcal{R}_r^{(n)}\Big(\widehat{F}_{(\widehat{\theta},\widehat{h})},F\Big) \leq c_{r,1}\bigg\|\frac{\ln(n)}{nh^*_{\mathcal{K},f}}\bigg\|_{\frac{r}{2}}^{\frac{1}{2}} + c_{r,2}n^{-\frac{1}{2}}.$$

2.1.4. *Extension to the case $\theta^*\notin\mathbb{S}^1$.* Define $f_{\theta^*}(t) = f\big(|\theta^*|_2 t\big)$, $\vartheta^* = \theta^*/|\theta^*|_2$ and let $F_{\theta^*}(x) := f_{\theta^*}\big(x^\top\vartheta^*\big)$. Obviously,

$$f_{\theta^*}\big(x^\top\vartheta^*\big) = f\big(x^\top\theta^*\big),\ \forall x \quad\Rightarrow\quad F_{\theta^*}(\cdot)\equiv F(\cdot)$$

so the estimation of $F(\cdot)$ is equivalent to the estimation of $F_{\theta^*}(\cdot)$. Moreover $\vartheta\in\mathbb{S}^1$ and, therefore, results obtained in Theorems 1 and 2 are applicable. To do that it suffices to replace f by f_{θ^*} in the definition of $h^*_{\mathcal{K},f}(\cdot)$. We note, however, that there is no any general receipt of expressing $h^*_{\mathcal{K},f_{\theta^*}}(\cdot)$ via $h^*_{\mathcal{K},f}(\cdot)$, although in particular cases (mainly in adaptive estimation over the collection of classes of smooth functions) it is often possible.

2.2. *Adaptive estimation.* In this section we first apply the local oracle inequality given in Theorem 1 to the problem of pointwise adaptive estimation over a collection of Hölder classes. Next, we study the problem of adaptive estimation under the L_r losses over a collection of Nikol'skii classes. The corresponding results are deduced from the global oracle inequality proved in Theorem 2.

Throughout this section we will assume that the kernel $\mathcal{K}$ obeys additionally Assumption 5 below. Introduce the following notation: for any $a>0$ let $m_a\in\mathbb{N}$ be the maximal integer strictly less than a.

ASSUMPTION 5. *There exists $\boldsymbol{b}>0$ such that*

$$\int z^j K(z)\mathrm{d}z = 0,\ \ \forall j=1,\dots,m_{\boldsymbol{b}}.$$

2.2.1. *Pointwise adaptive estimation.* We start this section with the definition of the Hölder class of functions.

DEFINITION 1. *Let $\beta > 0$ and $L > 0$. A function $\ell : \mathbb{R} \to \mathbb{R}$ belongs to the Hölder class $\mathbb{H}(\beta, L)$ if ℓ is m_β-times continuously differentiable, $\|\ell^{(m)}\|_\infty \le L$, $\forall m \le m_\beta$, and*

$$\left|\ell^{(m_\beta)}(t+h) - \ell^{(m_\beta)}(t)\right| \le L h^{\beta - m_\beta}, \quad \forall t \in \mathbb{R}, h > 0.$$

The aim is to estimate the function $F(x)$ at a given point $x \in [-1/2, 1/2]^2$ under the assumption that $F \in \mathbb{F}(\boldsymbol{b}) := \bigcup_{\beta \le \boldsymbol{b}} \bigcup_{L>0} \mathbb{F}_2(\beta, L)$, where

$$\mathbb{F}_d(\beta, L) = \left\{ F : \mathbb{R}^d \to \mathbb{R} \mid F(z) = f\left(z^\top \theta\right), \ f \in \mathbb{H}(\beta, L), \ \theta \in \mathbb{S}^{d-1} \right\},$$

the constant $\boldsymbol{b}$ is from Assumption 5, and $d \ge 2$ is the dimension. We will see that $\boldsymbol{b}$ can be an arbitrary number but it must be chosen *a priory*.

THEOREM 3. *Let $\boldsymbol{b} > 0$ be fixed and let Assumptions 4 and 5 hold. Then, for any $\beta \le \boldsymbol{b}$, $L > 0$ and $x \in [-1/2, 1/2]^2$,*

$$\sup_{F \in \mathbb{F}_2(\beta, L)} \mathcal{R}_{r,x}^{(n)}\left(\widehat{F}_{(\widehat{\theta}, \widehat{h})}, F\right) \le \varkappa_1 \psi_n(\beta, L),$$

where $\psi_n(\beta, L) = L^{\frac{1}{2\beta+1}} \left(n^{-1}\ln(n)\right)^{\frac{\beta}{2\beta+1}}$ and $\varkappa_1$ is independent of n.

The proof of the theorem it is based on the evaluation of the uniform, over $\mathbb{H}_d(\beta, L)$, lower bound for $h^*_{\mathcal{K},f}(\cdot)$ and on the application of Theorem 1. We note that a similar upper bound for the minimax risk was established in Goldenshluger and Lepski (2008) in the framework of Gaussian white noise model, but the estimation procedure used there is different from our selection rule.

The main question, however, is whether $\psi_n(\beta, L)$ coincides with the minimax rate of convergence for any given value of β and L? To answer this question we will need some additional assumptions on the density of the noise variable ε_1 and design variable X_1.

ASSUMPTION 6. *There exist $\mathfrak{q}, \mathfrak{Q} > 0$ such that for any $\upsilon_1, \upsilon_2 \in [-\mathfrak{q}, \mathfrak{q}]$*

$$\int_{\mathbb{R}} \frac{p(y+\upsilon_1)p(y+\upsilon_2)}{p(y)} \mathrm{d}y \le 1 + \mathfrak{Q}\left|\upsilon_1 \upsilon_2\right|.$$

It is easy to see that the density of the normal law $\mathcal{N}(0,\sigma^2),\ \sigma^2>0$, obeys the aforementioned assumption. In general, this assumption is fulfilled, if the density p is regular and decrease rapidly at infinity. More precisely, if the Fisher information corresponding to the density p is finite and the function $\int\big[p'(y+\cdot)\big]^2 p^{-1}(y)\mathrm{d}y$ is continuous at zero then Assumption 6 is verified.

ASSUMPTION 7. *There exist $\mathfrak{g}>0$ and $\varpi>1$ such that*

$$g(x)\le\frac{\mathfrak{g}}{1+|x|_2^{\varpi}},\quad \forall x\in\mathbb{R}^d,$$

where $|\cdot|_2$ is the Euclidian vector norm on $\mathbb{R}^d$.

We remark that imposed assumption is very weak and it is checked for the majority of probability distributions emerging in statistics.

THEOREM 4. *Let Assumptions 6 and 7 be fulfilled. Then for any $x\in[-1/2,1/2]^d$, $d\ge2$, $r\ge1$, $\beta,L>0$, and any $n\in\mathbb{N}^*$ large enough,*

$$\inf_{\widetilde{F}}\sup_{F\in\mathbb{F}_d(\beta,L)}\mathcal{R}_{r,x}^{(n)}\left(\widetilde{F},F\right)\ge\varkappa_2\psi_n(\beta,L),$$

where the infimum is over all possible estimators. Here $\varkappa_2$ is a numerical constant independent of n and L.

To the best of our knowledge this lower bound is new. We would like to emphasize that Assumption 6 under which this theorem is proved is close to be necessary. It is not difficult to provide examples in which this condition does not hold and the assertion of Theorem 4 is not true anymore.

We conclude from Theorems 3 and 4 that the estimator $\widehat{F}_{(\widehat{\theta},\widehat{h})}$ is minimax adaptive with respect to the collection of classes $\big\{\mathbb{F}_d(\beta,L),\ \ \beta\le\boldsymbol{b},\ L>0\big\}$. As already mentioned, this result is quite surprising. Indeed, if for example the directional vector $\theta=(1,0)^\top$, i.e. is known, then $\mathbb{F}(\beta,L)=\mathbb{H}(\beta,L)$, and the considered estimation problem can be easily reduced to estimation of f at a given point in the univariate regression model. As it is shown in Gaïffas (2007) the adaptive estimator over the collection $\big\{\mathbb{H}(\beta,L),\ \ \beta\le\boldsymbol{b},\ L>0\big\}$ does not exist and the prise to be paid for adaption appears. The latter means that the asymptotic bound on the minimax risk provided by adaptive estimator differs from the minimax rate of convergence by some factor. This factor for the majority of known results is $\ln(n)$.

Also, we would like to emphasize that the assertion of Theorem 4 is proved for arbitrary dimension.

2.2.2. *Adaptive estimation under the L_r losses.* We start this section with the definition of the Nikol'skii class of functions.

DEFINITION 2. *Let $\beta>0$, $L>0$ and $p\in[1,\infty)$ be fixed. A function $\ell:\mathbb{R}\to\mathbb{R}$ belongs to the Nikol'skii class $\mathbb{N}_p(\beta,L)$ if ℓ is m_β-times continuously differentiable and*

$$\left(\int_{\mathbb{R}}\left|\ell^{(m)}(t)\right|^p\mathrm{d}t\right)^{\frac{1}{p}}\leq L,\ \forall m=0,\dots,m_\beta;$$
$$\left(\int_{\mathbb{R}}\left|\ell^{(m_\beta)}(t+h)-\ell^{(m_\beta)}(t)\right|^p\mathrm{d}z\right)^{\frac{1}{p}}\leq Lh^{\beta-m_\beta},\ \forall h>0.$$

Later on we assume that $\mathbb{N}_p(\beta,L)=\mathbb{H}(\beta,L)$ if $p=\infty$.

Here the target of estimation is the entire function $F(\cdot)$ under the assumption that $F\in\mathbb{F}_p(\boldsymbol{b}):=\bigcup_{\beta\leq\boldsymbol{b}}\bigcup_{L>0}\mathbb{F}_{2,p}(\beta,L)$, where

$$\mathbb{F}_{d,p}(\beta,L)=\left\{F:\mathbb{R}^d\to\mathbb{R}\mid F(z)=f\big(z^\top\theta\big),\ f\in\mathbb{N}_p(\beta,L),\ \theta\in\mathbb{S}^{d-1}\right\}.$$

Let us briefly discuss the applicability of Theorem 2 which requires that $f\in\mathbb{F}(\beta_0,M)$. In order to guarantee it we will assume that $\beta p>1$. The latter assumption is standard in estimation of functions possessing inhomogeneous smoothness, see for example, Donoho et al. (1995), Lepski et al. (1997), Kerkyacharian et al. (2008). If $\beta p>1$ the embedding $\mathbb{N}_p(\beta,L)\subset\mathbb{H}(\beta-1/p,cL)$ with some absolute constant $c>0$ guarantees that $f\in\mathbb{F}(\beta_0,M)$ with $\beta_0=\beta-1/p, M=cL$ so Theorem 2 is applicable.

THEOREM 5. *Let $\boldsymbol{b}>0$ be fixed, and let Assumptions 4 and 5 hold. Then, for any $L>0$, $p>1$, $p^{-1}<\beta\leq\boldsymbol{b}$ and $r\geq1$,*

$$\sup_{F\in\mathbb{F}_{2,p}(\beta,L)}\mathcal{R}_r^{(n)}\Big(\widehat{F}_{(\widehat{\theta},\widehat{h})},F\Big)\leq\varkappa_3\varphi_n(\beta,L,p),$$

where $\varkappa_3$ is independent of n and

$$\varphi_n(\beta,L,p)=\begin{cases}L^{1/(2\beta+1)}\left(n^{-1}\ln(n)\right)^{\frac{\beta}{2\beta+1}}, & (2\beta+1)p>r;\\ L^{1/(2\beta+1)}\left(n^{-1}\ln(n)\right)^{\frac{\beta}{2\beta+1}}\left[\ln(n)\right]^{\frac{1}{r}}, & (2\beta+1)p=r;\\ L^{\frac{1/2-1/r}{\beta-1/p+1/2}}\left(n^{-1}\ln(n)\right)^{\frac{\beta-1/p+1/r}{2\beta-2/p+1}}, & (2\beta+1)p<r.\end{cases}$$

Note that $\mathbb{F}_{2,p}(\beta,L)\supset\mathbb{N}_p(\beta,L)$. Indeed, the class $\mathbb{N}_p(\beta,L)$ can be viewed as a class of functions F satisfying $F(\cdot)=f(\theta^\top\cdot)$ with $\theta=(1,0)^\top$. Then,

the problem of estimating such (2-variate) functions can be reduced to the estimation of univariate functions observed in the one-dimensional regression model.

There are at least two observations appearing in view of the latter remark. First, the upper bound given in Theorem 5 generalizes the results obtained in the univariate regression, see for instance, Donoho et al. (1995), Delyon and Juditsky (1996), Baraud (2002), Kerkyacharian and Picard (2004), Kulik and Raimondo (2009), Zhang et al. (2002) in several directions. In particular, the majority of the papers treats the Gaussian errors or the errors possessing exponential moment. The exception is the paper Baraud (2002) in which some part of results is obtained under very weak assumption imposed on the noise (weaker than our Assumption 1). However, these results are available only if $p = r = 2$.

Next, the rate of convergence for the latter problem (which can be found in Chesneau (2007)) is also the lower bound for the minimax risk defined on $\mathbb{F}_{2,p}(\beta, L)$. Under assumption $\beta p > 1$ this rate of convergence is given by

$$\phi_n(\beta, L, p) = \begin{cases} L^{1/(2\beta+1)} n^{-\frac{\beta}{2\beta+1}}, & (2\beta+1)p > r; \\ L^{1/(2\beta+1)} \left(n^{-1}\ln(n)\right)^{\frac{\beta}{2\beta+1}}, & (2\beta+1)p = r; \\ L^{\frac{1/2-1/r}{\beta-1/p+1/2}} \left(n^{-1}\ln(n)\right)^{\frac{\beta-1/p+1/r}{2\beta-2/p+1}}, & (2\beta+1)p < r. \end{cases}$$

The minimax rate of convergence in the case $(2\beta+1)p = r$ remains an open problem, and the rate presented in the middle line above is only the lower asymptotic bound for the minimax risk.

Thus the proposed estimator $\widehat{F}_{(\widehat{\theta},\widehat{h})}$ is adaptive whenever $(2\beta+1)p < r$. In the case $(2\beta+1)p \geq r$ we loose only a logarithmic factor with respect to the optimal rate and, as mentioned in Introduction, the construction of an adaptive estimator over the collection $\{\mathbb{F}_{2,p}(\beta, L),\ \beta > 0,\ L > 0\}$ in this case remains an open problem. In view of the latter remark we conjecture that the presented lower bound is correct and, therefore, the upper bound result has to be improved.

3. Proofs. We start this section with presenting the quantities used in the description of the selection rule led to the adaptive estimator $\widehat{F}_{(\widehat{\theta},\widehat{h})}$.

3.1. *Important quantities.* Set

$$
\begin{aligned}
c_1(n) &= 730\ln\Big(16n^2\underline{g}^{-\frac{1}{2}}\big[12Q+\sqrt{2}\big]\Big)+8r\ln(n)+394;\\
c_2(n) &= 730\ln\Big(16n^2\tau\underline{g}^{-\frac{1}{2}}\big[12Q+\sqrt{2}\big]\Big)+8r\ln(n)+394;\\
c_3(n) &= 365\ln\Big(2n^2Q\underline{g}^{-\frac{1}{2}}\Big)+8r\ln(n)+197;\\
c_4(n) &= 365\ln\Big(2n^2\tau Q\underline{g}^{-\frac{1}{2}}\Big)+8r\ln(n)+197,
\end{aligned}
$$

and define

$$
\begin{aligned}
C_1(n) &= 2\sqrt{2}\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_1(n)}+(8/3)c_1(n)\big(\ln(n)\big)^{-\frac{2+\omega}{2\omega}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2;\\
C_2(n) &= 2\sqrt{2}(\sigma\vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_2(n)}\\
&\quad +(8/3)c_2(n)\big(\ln(n)\big)^{-\frac{1}{2}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2\big(\Omega^{-1}(4r+1)\big)^{\frac{1}{\omega}};\\
C_3(n) &= 2\sqrt{2}\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_3(n)}+(8/3)\underline{g}^{-1}\|\mathcal{K}\|_\infty^2c_3(n)\big(n\mathfrak{h}^2\big)^{-\frac{1}{2}};\\
C_4(n) &= 2\sqrt{2}(\sigma\vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_4(n)}+(8/3)\tau c_4(n)\big(n\mathfrak{h}^2\big)^{-\frac{1}{2}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2;\\
C_5(n) &= \|\mathcal{K}\|_1^2+\big(n\mathfrak{h}^2\big)^{-\frac{1}{2}}C_4(n),
\end{aligned}
$$

where, remind $\mathfrak{h}$ is given in (1.5), and

$$
\sigma^2=\sup_{p\in\mathfrak{P}}\int_{\mathbb{R}}x^2p(x)\mathrm{d}x,\qquad \tau=\big(\Omega^{-1}(4r+1)\ln(n)\big)^{\frac{1}{\omega}}.
$$

In spite of the cumbersome expressions, it is easy to see that

$$
\sup_{n\geq 3}\frac{C_i(n)}{\sqrt{\ln(n)}}=:C_i<\infty,\;\; i=1,2,\qquad \sup_{n\geq 3}C_5(n)=:C_5<\infty. \tag{3.1}
$$

3.2. *Proof of Theorem 1.* We start the proof with establishing technical lemmas whose proofs are postponed to Appendix.

3.2.1. *Auxiliary results.* For any $\theta,\nu\in\mathbb{S}^1$ and $h\in\big[2^{-1}h_{\min},1\big]$ denote

$$
\begin{aligned}
S_{(\theta,h)(\nu,h)}(x) &= \det\big(E_{(\theta,h)(\nu,h)}\big)\int K(E_{(\theta,h)(\nu,h)}(t-x))F(t)\mathrm{d}t,\\
S_{(\theta,h)}(x) &= \det\big(E_{(\theta,h)}\big)\int K(E_{(\theta,h)}(t-x))F(t)\mathrm{d}t.
\end{aligned}
$$

For ease of notation, we write $h^*_f=h^*_{\mathcal{K},f}(x^\top\theta^*)$.

LEMMA 1. *Grant Assumption 4. Then, for any $\nu \in \mathbb{S}^1$ and any bandwidths $\eta, h \in [2^{-1}h_{\min}, 1]$ satisfying $\eta \le h \le 2^{-1}h_f^*$, one has*

$$\left|S_{(\theta^*,h)(\nu,h)}(x) - S_{(\nu,h)}(x)\right| \le 2(h_f^*)^{-1/2}\|\mathcal{K}\|_\infty^2\sqrt{n^{-1}\ln(n)};$$
$$\left|S_{(\nu,h)}(x) - S_{(\nu,\eta)}(x)\right| \le 2(h_f^*)^{-1/2}\|\mathcal{K}\|_\infty^2\sqrt{n^{-1}\ln(n)};$$
$$\left|S_{(\theta^*,h)} - F(x)\right| \le (h_f^*)^{-1/2}\|\mathcal{K}\|_\infty\sqrt{n^{-1}\ln(n)}.$$

Let $\mathcal{E}_{a,A}$ with $a \in (0,1], A \ge 1$, be a set of 2×2 matrices satisfying

$$\left|\det(E)\right| \le A, \quad |E|_\infty \le (\sqrt{2a})^{-1}\left|\det(E)\right|. \tag{3.2}$$

Here $|E|_\infty = \max_{i,j}|E_{i,j}|$ denotes the matrix supremum norm. Set for any $E \in \mathcal{E}_{a,A}$

$$J(x,E) \;=\; \sqrt{\left|\det(E)\right|}K\big(E(x-t)\big)[g(x)]^{-1}, \quad x \in \mathbb{R}^2;$$

and consider the following random fields defined on $\mathcal{E}_{a,A}$:

$$\eta_{n,t}(E) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{J(X_i,E)F(X_i) - \mathbb{E}_X^{(n)}\left[J(X_i,E)F(X_i)\right]\right\},$$
$$\xi_{n,t}(E) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}J(X_i,E)\varepsilon_i.$$

Denote finally by $\mathcal{E}_*$ the set of matrices $\mathcal{E}_{a,A}$ with $a = 1/8$ and $A = (h_{\min})^{-1}$.

LEMMA 2. *Grant Assumptions 1-4. Then, for any $n \ge 3$ and any $r \ge 1$,*

$$\mathbb{P}_{X,\varepsilon}^{(n)}\left\{\sup_{E\in\mathcal{E}_*}\Big[\left|\eta_{n,t}(E)\right| + \left|\xi_{n,t}(E)\right|\Big] \ge C_1(n)\|F\|_\infty + C_2(n)\right\} \le (8+\Upsilon)n^{-4r}.$$

The expressions $C_1(n)$ and $C_2(n)$ are given in Section 3.1.

LEMMA 3. *Grant Assumptions 1-4. Then, for any $n \ge n_0 \vee n_1$,*

$$\sup_{\theta^*\in\mathbb{S}^1}\sup_{f\in\mathbb{F}(\beta_0,M)}\mathbb{P}_F^{(n)}\left\{\widehat{F}_\infty \notin \Big[\|F\|_\infty, 3M + 4C_5(n)\Big]\right\} \le (8+\Upsilon)n^{-4r}.$$

The numbers n_0, n_1 are defined in (1.6) and $C_5(n)$ is defined in Section 3.1.

3.2.2. *Proof of Theorem 1.* In view of Jensen's inequality an upper bound for $\mathcal{R}^{(n)}_{r,x}$, $r \geq 2$, will suffice to complete the proof.

Let $h^* \in \mathcal{H}_n$ be such that $2h^* \leq h^*_f < 4h^*$. Introduce the random events

$$\mathcal{A} = \left\{ R^{(1)}_x(\theta^*, h^*) + R^{(2)}_x(h^*) = 0 \right\}, \quad \mathcal{B} = \left\{ \widehat{F}_\infty \in \Big[\|F\|_\infty, 3M + 4C_5(n) \Big] \right\},$$

and let $\overline{\mathcal{A}}$ and $\overline{\mathcal{B}}$ denote events complimentary to $\mathcal{A}$ and $\mathcal{B}$ respectively. We split the proof into three steps.

Risk computation under $\mathcal{A} \cap \mathcal{B}$. First, the following inclusion holds:

$$\mathcal{A} \subseteq \big\{ \widehat{h} \geq h^* \big\}. \tag{3.3}$$

Indeed, in view of the definition of the couple $(\widehat{\theta}, \widehat{h})$ we have

$$\begin{aligned} 1_{\mathcal{A}}\, \mathrm{TH}(h^*) &= 1_{\mathcal{A}} \left\{ R^{(1)}_x(\theta^*, h^*) + R^{(2)}_x(h^*) + \mathrm{TH}(h^*) \right\} \\ &\geq 1_{\mathcal{A}} \left\{ R^{(1)}_x(\widehat{\theta}, \widehat{h}) + R^{(2)}_x(\widehat{h}) + \mathrm{TH}(\widehat{h}) \right\} \geq 1_{\mathcal{A}}\, \mathrm{TH}(\widehat{h}). \end{aligned}$$

It remains to note that the mapping $\eta \mapsto \mathrm{TH}(\eta)$ is decreasing so inclusion (3.3) follows. Next, the triangle inequality yields

$$\begin{aligned} \left| \widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - F(x) \right| \;\leq\; & \left| \widehat{F}_{(\theta^*,h^*)}(x) - F(x) \right| + \left| \widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - \widehat{F}_{(\widehat{\theta},h^*)}(x) \right| \\ & + \left| \widehat{F}_{(\theta^*,h^*)(\widehat{\theta},h^*)}(x) - \widehat{F}_{(\widehat{\theta},h^*)}(x) \right| \\ & + \left| \widehat{F}_{(\theta^*,h^*)(\widehat{\theta},h^*)}(x) - \widehat{F}_{(\theta^*,h^*)}(x) \right|. \end{aligned} \tag{3.4}$$

1^0. We have in view of (3.3) and the definition of $R^{(2)}_x$ that

$$1_{\mathcal{A}} \left| \widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - \widehat{F}_{(\widehat{\theta},h^*)}(x) \right| \leq 1_{\mathcal{A}} \big[R^{(2)}_x(\widehat{h}) + \mathrm{TH}(h^*) \big]. \tag{3.5}$$

The definition of $R^{(1)}_x(\cdot,\cdot)$ implies

$$\begin{aligned} 1_{\mathcal{A}} \left| \widehat{F}_{(\theta^*,h^*)(\widehat{\theta},h^*)}(x) - \widehat{F}_{(\widehat{\theta},h^*)}(x) \right| &\leq 1_{\mathcal{A}} \big[R^{(1)}_x(\theta^*, h^*) + \mathrm{TH}(h^*) \big] \\ &= 1_{\mathcal{A}}\, \mathrm{TH}(h^*). \end{aligned} \tag{3.6}$$

Note that $E_{(\theta,h)(\nu,h)} = \pm E_{(\nu,h)(\theta,h)}$ for any θ, ν and h. Hence,

$$\widehat{F}_{(\theta^*,h^*)(\widehat{\theta},h^*)}(\cdot) \equiv \widehat{F}_{(\widehat{\theta},h^*)(\theta^*,h^*)}(\cdot)$$

since $\mathcal{K}$ is symmetric. It yields together with (3.3) and the definition of $R_x^{(1)}$

$$1_{\mathcal{A}}\left|\widehat{F}_{(\theta^*,h^*)(\widehat{\theta},h^*)}(x)-\widehat{F}_{(\theta^*,h^*)}(x)\right| = 1_{\mathcal{A}}\left|\widehat{F}_{(\widehat{\theta},h^*)(\theta^*,h^*)}(x)-\widehat{F}_{(\theta^*,h^*)}(x)\right| \tag{3.7}$$
$$\leq 1_{\mathcal{A}}\big[R_x^{(1)}(\widehat{\theta},\widehat{h})+\mathrm{TH}(h^*)\big].$$

We obtain from (3.4), (3.5), (3.6) and (3.7) that

$$1_{\mathcal{A}}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x)-F(x)\right| \leq 1_{\mathcal{A}}\big[R_x^{(1)}(\widehat{\theta},\widehat{h})+R_x^{(2)}(\widehat{h})\big]+3\,\mathrm{TH}(h^*) + \left|\widehat{F}_{(\theta^*,h^*)}(x)-F(x)\right|.$$

Noting that in view of the definition of $(\widehat{\theta},\widehat{h})$

$$\begin{aligned} R_x^{(1)}(\widehat{\theta},\widehat{h})+R_x^{(2)}(\widehat{h}) &\leq R_x^{(1)}(\widehat{\theta},\widehat{h})+R_x^{(2)}(\widehat{h})+\mathrm{TH}(\widehat{h}) \\ &\leq R_x^{(1)}(\theta^*,h^*)+R_x^{(2)}(h^*)+\mathrm{TH}(h^*), \end{aligned}$$

we obtain

$$1_{\mathcal{A}}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x)-F(x)\right| \leq 4\,\mathrm{TH}(h^*)+\left|\widehat{F}_{(\theta^*,h^*)}(x)-F(x)\right|. \tag{3.8}$$

Note also that for any $\eta\in\mathcal{H}_n$

$$\begin{aligned} 1_{\mathcal{B}}\,\mathrm{TH}(\eta) &\leq 2\Big[\|\mathcal{K}\|_\infty^2\sqrt{\ln(n)}+(3M+4C_5)C_1(n)+C_2(n)\Big](\eta n)^{-\frac{1}{2}} \\ &\leq C_6\sqrt{(\eta n)^{-1}\ln(n)}, \end{aligned}$$

where $C_6=2\|\mathcal{K}\|_\infty^2+2(3M+4C_5)C_1+2C_2$ and C_1,C_2 and C_5 are defined in (3.1). Since $\mathrm{TH}(h^*)\leq\mathrm{TH}(h_f^*/4)$, this bound and (3.8) yield

$$1_{\mathcal{A}\cap\mathcal{B}}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x)-F(x)\right|\leq 8C_6\sqrt{\frac{\ln(n)}{nh_f^*}}+\left|\widehat{F}_{(\theta^*,h^*)}(x)-F(x)\right|. \tag{3.9}$$

2^0. Note that $E_{(\theta,h)(\nu,h)},E_{(\theta,h)}\in\mathcal{E}_*$ for any $\theta,\nu\in\mathbb{S}^1$, $h\in[h_{\min},1]$. Set

$$\widehat{F}(E,x)=\det(E)\sum_{i=1}^n K\big(E(X_i-x)\big)g^{-1}(X_i)Y_i,\quad E\in\mathcal{E}_*.$$

The following "approximation + stochastic part" decomposition of $\widehat{F}(E,x)$ will be useful in the sequel:

$$\widehat{F}(E,x) = \det(E)\int K(E(t-x))F(t)\mathrm{d}t \tag{3.10}$$
$$+ \sqrt{\det(E)/n}\big[\eta_{n,t}(E)+\xi_{n,t}(E)\big],$$

where $\eta_{n,t}(E)$ and $\xi_{n,t}(E)$ are defined before the statement of Lemma 2. Thus, we have

$$\left|\widehat{F}_{(\theta^*,h^*)}(x) - F(x)\right| \leq \left|S_{(\theta^*,h^*)} - F(x)\right|$$
$$+\big(n^{-1}\det\big(E_{(\theta^*,h^*)}\big)^{\frac{1}{2}}\big[\eta_{n,t}\big(E_{(\theta^*,h^*)}\big) + \xi_{n,t}\big((E_{(\theta^*,h^*)})\big].$$

Taking into account that $\det\big(E_{(\theta^*,h^*)}\big) = (h^*)^{-1} \leq 4(h^*_f)^{-1}$ in view of the definition of h^* and using the third assertion of Lemma 1 we obtain

$$\left|\widehat{F}_{(\theta^*,h^*)}(x) - F(x)\right|$$
$$\leq \big[(nh^*_f)^{\frac{1}{2}}\big]\Big[\sqrt{\ln(n)}\|\mathcal{K}\|_\infty + 2\big|\eta_{n,t}\big(E_{(\theta^*,h^*)}\big) + \xi_{n,t}\big(E_{(\theta^*,h^*)}\big)\big|\Big]$$

Applying the Rosenthal inequality to $\eta_{n,t}\big(E_{(\theta^*,h^*)}\big) + \xi_{n,t}\big(E_{(\theta^*,h^*)}\big)$ which is a sum of centered independent random variables we obtain from (3.9)

$$\left\{\mathbb{E}_F^{(n)}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - F(x)\right|^r 1_{\mathcal{A}\cap\mathcal{B}}\right\}^{\frac{1}{r}} \leq \widetilde{c}_0\sqrt{(nh^*_f)^{-1}\ln(n)}, \tag{3.11}$$

where $\widetilde{c}_0$ is independent of F and n.

Risk computation under $\overline{\mathcal{B}}$. Since $f \in \mathbb{F}(\beta_0, M)$ we have the following obvious bound

$$\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - F(x)\right| \leq n\Big\{M\big(1 + \underline{g}^{-1}\|K\|_\infty\big) + \underline{g}^{-1}n^{-1}\|\mathcal{K}\|^2_\infty\sum_{i=1}|\varepsilon_i|\Big\},$$

where we have also used that $nh_{\min} > 1$. Hence, in view of the Rosenthal inequality we obtain

$$\left[\mathbb{E}_F^{(n)}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - F(x)\right|^{2r}\right]^{\frac{1}{2r}} \leq n\widetilde{c}_1,$$

where $\widetilde{c}_1$ is independent of F and n.

The use of the Cauchy-Schwartz inequality together with the statement of Lemma 3 lead to the following bound:

$$\left\{\mathbb{E}_F^{(n)}\left|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x) - F(x)\right|^r 1_{\overline{\mathcal{B}}}\right\}^{\frac{1}{r}} \leq n\widetilde{c}_1\left[\mathbb{P}_F^{(n)}(\overline{\mathcal{B}})\right]^{\frac{1}{2r}} \leq \widetilde{c}_1(8+\Upsilon)^{\frac{1}{2r}}n^{-1}. \tag{3.12}$$

Risk computation under $\overline{\mathcal{A}}\cap\mathcal{B}$. We note that

$$\mathbb{P}_F^{(n)}\big(\overline{\mathcal{A}}\cap\mathcal{B}\big)\le\mathbb{P}_F^{(n)}\Big\{R_x^{(1)}(\theta^*,h^*)>0,\ \mathcal{B}\Big\}+\mathbb{P}_F^{(n)}\Big\{R_x^{(2)}(h^*)>0,\ \mathcal{B}\Big\}.$$

1^0. First, let us bound from above $\mathbb{P}_F^{(n)}\Big\{R_x^{(1)}(\theta^*,h^*)>0,\mathcal{B}\Big\}$. We have

$$\Big\{R_x^{(1)}(\theta^*,h^*)>0\Big\}=\bigcup_{\eta\in\mathcal{H}_n,\eta\le h^*}\Big\{\sup_{\nu\in\mathbb{S}^1}\big|\widehat{F}_{(\theta^*,\eta)(\nu,\eta)}(x)-\widehat{F}_{(\nu,\eta)}(x)\big|>\mathrm{TH}\left(\eta\right)\Big\}$$

and, therefore

$$\mathbb{P}_F^{(n)}\Big\{R_x^{(1)}(\theta^*,h^*)>0,\ \mathcal{B}\Big\}\le\sum_{k:\,2^{-1}h_{\min}\le 2^{-k}\le h^*}\mathbb{P}_F^{(n)}\left\{\sup_{\nu\in\mathbb{S}^1}\big|\widehat{F}_{(\theta^*,2^{-k})(\nu,2^{-k})}(x)-\widehat{F}_{(\nu,2^{-k})}(x)\big|>\mathrm{TH}\left(2^{-k}\right),\ \mathcal{B}\right\}. \tag{3.13}$$

Thus, denoting by $\varsigma_n=\sup_{E\in\mathcal{E}_*}\big[\big|\eta_{n,t}(E)\big|+\big|\xi_{n,t}(E)\big|\big]$ and using (3.10) together with the first assertion of Lemma 1 we obtain for any $k:2^{-k}\le h^*$

$$\begin{aligned}&\sup_{\nu\in\mathbb{S}^1}\Big|\widehat{F}_{(\theta^*,2^{-k})(\nu,2^{-k})}(x)-\widehat{F}_{(\nu,2^{-k})}(x)\Big|\\&\le 2(h_f^*)^{-1/2}\|\mathcal{K}\|_\infty^2\sqrt{n^{-1}\ln(n)}+2\sqrt{2^k}n^{-1/2}\,\varsigma_n\\&\le 2\,\|\mathcal{K}\|_\infty^2\sqrt{2^kn^{-1}\ln(n)}+2\big(2^kn^{-1}\big)^{1/2}\,\varsigma_n.\end{aligned}\tag{3.14}$$

Here we have also used that $2^{-1}h_f^*\ge 2^{-k}$. Note also that

$$1_{\mathcal{B}}\,\mathrm{TH}(\eta)\ge 2\|\mathcal{K}\|_\infty^2\sqrt{\frac{\ln(n)}{\eta n}}+\frac{2}{\sqrt{\eta n}}\Big\{\|F\|_\infty C_1(n)+C_2(n)\Big\},\tag{3.15}$$

and, therefore, we obtain from (3.14) for any k satisfying $2^{-k}\le h^*$

$$\begin{aligned}&\mathbb{P}_F^{(n)}\left\{\sup_{\nu\in\mathbb{S}^1}\big|\widehat{F}_{(\theta^*,2^{-k})(\nu,2^{-k})}(x)-\widehat{F}_{(\nu,2^{-k})}(x)\big|>\mathrm{TH}\left(2^{-k}\right),\ \mathcal{B}\right\}\\&\le\mathbb{P}_{X,\varepsilon}^{(n)}\Big\{\varsigma_n\ge\|F\|_\infty C_1(n)+C_2(n)\Big\}\le(8+\Upsilon)n^{-4r},\end{aligned}$$

in view of Lemma 2. It yields, together with (3.13)

$$\mathbb{P}_F^{(n)}\Big(R_x^{(1)}(\theta^*,h^*)>0,\ \mathcal{B}\Big)\le(8+\Upsilon)\log_2(n)n^{-4r}.\tag{3.16}$$

2^0. Now, let us bound from above $\mathbb{P}_F^{(n)}\Big\{R_x^{(2)}(h^*)>0,\ \mathcal{B}\Big\}$. We have

$$\Big\{R_x^{(2)}(h^*)>0\Big\}=\bigcup_{\eta\in\mathcal{H}_n:\,\eta\le h^*\le h^*}\left\{\sup_{\theta\in\mathbb{S}^1}\Big|\widehat{F}_{(\theta,h^*)}(x)-\widehat{F}_{(\theta,\eta)}(x)\Big|>\mathrm{TH}\left(2^{-k}\right)\right\},$$

hence,

$$\mathbb{P}_F^{(n)}\Big\{R_x^{(2)}(h^*)>0,\ \mathcal{B}\Big\} \tag{3.17}$$
$$\le\sum_{k:\,2^{-1}h_{\min}\le 2^{-k}\le h^*}\mathbb{P}_F^{(n)}\left\{\sup_{\theta\in\mathbb{S}^1}\Big|\widehat{F}_{(\theta,h^*)}(x)-\widehat{F}_{(\theta,2^{-k})}(x)\Big|>\mathrm{TH}\left(2^{-k}\right),\ \mathcal{B}\right\}.$$

Using (3.10) and the second assertion of Lemma 1 we obtain for any k satisfying $2^{-k}\le h^*$

$$\begin{aligned}&\sup_{\theta\in\mathbb{S}^1}\Big|\widehat{F}_{(\theta,h^*)}(x)-\widehat{F}_{(\theta,2^{-k})}(x)\Big|\\ \le\ &2(h_f^*)^{-1/2}\|\mathcal{K}\|_\infty^2\sqrt{n^{-1}\ln(n)}+2\sqrt{2^k}n^{-1/2}\,\varsigma_n\\ \le\ &2\,\|\mathcal{K}\|_\infty^2\sqrt{2^kn^{-1}\ln(n)}+2\big(2^kn^{-1}\big)^{1/2}\,\varsigma_n.\end{aligned}$$

Applying Lemma 1, we get in view of (3.15) for any k satisfying $2^{-k}\le h^*$

$$\mathbb{P}_F^{(n)}\left\{\sup_{\theta\in\mathbb{S}^1}\Big|\widehat{F}_{(\theta,h^*)}(x)-\widehat{F}_{(\theta,2^{-k})}(x)\Big|>\mathrm{TH}\left(2^{-k}\right),\ \mathcal{B}\right\}$$
$$\le\mathbb{P}_{X,\varepsilon}^{(n)}\Big\{\varsigma_n\ge\|F\|_\infty C_1(n)+C_2(n)\Big\}\le(8+\Upsilon)n^{-4r}.$$

It yields, together with (3.17)

$$\mathbb{P}_F^{(n)}\Big\{R_x^{(2)}(h^*)>0,\ \mathcal{B}\Big\}\le(8+\Upsilon)\log_2(n)n^{-4r}. \tag{3.18}$$

Thus, we obtain from (3.16) and (3.18) that

$$\mathbb{P}_F^{(n)}\big(\overline{\mathcal{A}}\cap\mathcal{B}\big)\le(8+\Upsilon)2\log_2(n)n^{-4r}.$$

It yields together with (3.2.2) that

$$\left\{\mathbb{E}_F^{(n)}\Big|\widehat{F}_{(\widehat{\theta},\widehat{h})}(x)-F(x)\Big|^r 1_{\overline{\mathcal{A}}\cap\mathcal{B}}\right\}^{\frac1r}\le n\widetilde{c}_1\left[\mathbb{P}_F^{(n)}\big(\overline{\mathcal{A}}\cap\mathcal{B}\big)\right]^{\frac{1}{2r}}\le\widetilde{c}_2n^{-\frac12}, \tag{3.19}$$

where $\widetilde{c}_2$ is independent of F and n.

The assertion of the theorem follows now from (3.11), (3.12) and (3.19).

■

3.3. *Proof of Theorem 3.* Using the standard computation of the bias of kernel estimators, under Assumptions 4 and 5 we get for any $f \in \mathbb{H}(\beta, L)$ and any $z \in \mathbb{R}$

$$\Delta_{\mathcal{K},f}(h,z) \leq \frac{Lh^{\beta}2^{-\beta}\|\mathcal{K}\|_{\infty}}{(1+\beta)m_{\beta}!} \leq \|\mathcal{K}\|_{\infty}Lh^{\beta}.$$

Since the right-hand side of the latter inequality is independent of z, we have $\Delta^{*}_{\mathcal{K},f}(h,z) \leq \|\mathcal{K}\|_{\infty}Lh^{\beta}$. This implies $h^{*}_{\mathcal{K},f}(z) \geq \big((Ln)^{-1}\ln(n)\big)^{1/(2\beta+1)}$ for any $z \in \mathbb{R}$ so the assertion of the theorem follows from Theorem 1. ∎

3.4. *Proof of Theorem 4.* We start this section with an auxiliary result used in the proof of the second assertion of the theorem. It was established in Kerkyacharian et al. (2008), Corollary 2 of Proposition 5 and, for convenience, we formulate it as Lemma 4 below.

3.4.1. *Auxiliary result.* The result cited below concerns a lower bound for estimators of an arbitrary mapping in the framework of abstract statistical model. We will not present it in full generality and below a version reduced to the estimation at a given point is provided.

Let $\mathcal{F}$ be a non-empty class of functions and let $F:\mathbb{R}^d \to \mathbb{R}$ be an unknown function from model (1.1)–(1.2). The aim is to estimate the functional $F(t)$, $t \in [-1/2, 1/2]^d$.

Introduce the following notation. For any given $F, G \in \mathcal{F}$ set

$$Z(F,G) = \prod_{i=1}^{n}\left[\frac{p\big(Y_i - F(X_i)\big)}{p\big(Y_i - G(X_i)\big)}\right].$$

Lemma 4. *Assume that for any sufficiently large $n \geq 1$ there exist a positive integer N_n, $c > 1$ and functions $F_0, \dots, F_{N_n} \in \mathcal{F}$ such that:*

$$|F_i(t) - F_0(t)| = \lambda_n, \qquad \forall i = 1, \dots, N_n; \tag{3.20}$$

$$\mathbb{E}^{(n)}_{F_0}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} Z\big(F_j, F_0\big)\right)^2 \leq c. \tag{3.21}$$

Then, for $r \geq 1$ and any $t \in [-1/2, 1/2]^d$,

$$\inf_{\widetilde{F}}\sup_{F\in\mathcal{F}}\left(\mathbb{E}^{(n)}_{F}\big|\widetilde{F}(t) - F(t)\big|^{r}\right)^{\frac{1}{r}} \geq \frac{1}{2}\left[1 - \sqrt{\frac{c-1}{c+3}}\right]\lambda_n.$$

3.4.2. *Proof of Theorem 4.* The proof is based on the construction of $F_0,\ldots,F_{N_n}$ satisfying conditions (3.20)–(3.21) of Lemma 4.

1^0. Firstly, we construct $F_0,\ldots,F_{N_n}$ and verify (3.20). Let $w:\mathbb{R}\to\mathbb{R}$ be such that $\mathrm{supp}(w)\subset(-1/2,1/2)$, $w\in\mathbb{H}(\beta,1)$, $\|w\|_\infty<\infty$, and $w(0)\neq 0$. Put $h=\big(\mathfrak{a}(L^2n)^{-1}\ln(n)\big)^{\frac{1}{2\beta+1}}$, where $\mathfrak{a}>0$ will be chosen later. Define

$$f(z)=Lh^\beta w\big(zh^{-1}\big)\,,\quad z\in\mathbb{R}. \tag{3.22}$$

For $b>0$ put $N_n=n^b$ assuming without loss of generality that N_n is an integer. The value of b will be determined later in order to satisfy (3.21).

Let $\{\vartheta_i,\ i=1,\ldots,N_n\}\subset\mathbb{S}^{d-1}$ be defined as follows:

$$\vartheta_i=\big(\theta_i^{(1)},\theta_i^{(2)},0,\ldots,0\big)^\top,\quad \theta_i^{(1)}=\cos(i/N_n),\ \theta_i^{(2)}=\sin(i/N_n).$$

Finally we set

$$F_0\equiv 0\quad\text{and}\quad F_i(x)=f\big(\vartheta_i^\top(x-t)\big)\,,\ i=1,\ldots,N_n. \tag{3.23}$$

Obviously, f defined by (3.22) belongs to $\mathbb{H}(\beta,L)$ so all F_i are in the class $\mathcal{F}=\mathbb{F}_d(\beta,L)$. Moreover, for any $i=1,\ldots,N_n$

$$|F_i(t)-F_0(t)|=|w(0)|L^{\frac{1}{2\beta+1}}\left(\mathfrak{a}n^{-1}\ln(n)\right)^{\frac{\beta}{2\beta+1}}=|w(0)|\mathfrak{a}^{\frac{\beta}{2\beta+1}}\psi_n(\beta,L).$$

We see that (3.20) holds with $\lambda_n=|w(0)|\mathfrak{a}^{\frac{\beta}{2\beta+1}}\psi_n(\beta,L)$.

2^0. Note that

$$\begin{aligned}&\mathbb{E}^{(n)}_{F_0}\left[\frac{1}{N_n}\sum_{j=1}^{N_n}Z\big(F_j,F_0\big)\right]^2\\&\quad=\frac{1}{N_n^2}\sum_{j=1}^{N_n}\mathbb{E}^{(n)}_{F_0}\big[Z^2\big(F_j,F_0\big)\big]+\frac{1}{N_n^2}\sum_{\substack{j,k=1,\\ j\neq k}}^{N_n}\mathbb{E}^{(n)}_{F_0}\Big[Z\big(F_j,F_0\big)Z\big(F_k,F_0\big)\Big].\end{aligned}$$

We have

$$\mathbb{E}^{(n)}_{F_0}\big\{Z^2\big(F_j,F_0\big)\big\}=\left\{\int_{\mathbb{R}^{d+1}}\frac{p^2\big(y-F_j(x)\big)}{p(y)}g(x)\mathrm{d}x\mathrm{d}y\right\}^n;$$

$$\mathbb{E}^{(n)}_{F_0}\Big\{Z\big(F_j,F_0\big)Z\big(F_k,F_0\big)\Big\}=\left\{\int_{\mathbb{R}^{d+1}}\frac{p\big(y-F_j(x)\big)p\big(y-F_k(x)\big)}{p(y)}g(x)\mathrm{d}x\mathrm{d}y\right\}^n.$$

Since $\lim_{n\to\infty}\ \sup_{j=1,\dots,N_n}\|F_j\|_\infty=0$, we have in view of Assumptions 6 and 7 for all n large enough

$$\int_{\mathbb{R}^{d+1}}\bigg[\frac{p^2\big(y-F_j(x)\big)}{p(y)}\bigg]g(x)\mathrm{d}x\mathrm{d}y\leq 1+\mathfrak{Q}\int_{\mathbb{R}^d}F_j^2(x)g(x)\mathrm{d}x$$
$$\leq 1+\mathfrak{Q}\mathfrak{g}\int_{\mathbb{R}^d}F_j^2(x)\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x; \tag{3.24}$$

$$\int_{\mathbb{R}^{d+1}}\bigg[\frac{p\big(y-F_j(x)\big)p\big(y-F_k(x)\big)}{p(y)}\bigg]g(x)\mathrm{d}x\mathrm{d}y\leq 1+\mathfrak{Q}\int_{\mathbb{R}^d}\big|F_j(x)F_k(x)\big|g(x)\mathrm{d}x$$
$$\leq 1+\mathfrak{Q}\mathfrak{g}\int_{\mathbb{R}^d}\big|F_j(x)F_k(x)\big|\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x. \tag{3.25}$$

Set $\theta_{j\perp}=\big(-\sin(j/N_n),\cos(j/N_n)\big)^\top$ and $\vartheta_{j\perp}=\big(\theta_{j\perp}^\top,0,\dots,0\big)^\top\in\mathbb{S}^{d-1}$. Denote for all $j=1,\dots,N_n$ by $\Theta_j^\top$, the orthogonal matrix $\big(\vartheta_i,\vartheta_{i\perp},\mathbf{e}_3,\dots,\mathbf{e}_d\big)$, where $\mathbf{e}_s$, $s=3,\dots d$, are the canonical basis vectors in $\mathbb{R}^d$. Integration by substitution with $\Theta_i(x-t)=v$ gives

$$\begin{aligned}\int_{\mathbb{R}^d}F_j^2(x)\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x &= L^2h^{2\beta}\int_{\mathbb{R}^d}w^2(v_1/h)\big(1+|x+\Theta_i^\top v|_2^{\varpi}\big)^{-1}\mathrm{d}v\\ &\leq C_\varpi L^2\|w\|_2^2h^{2\beta+1}=\mathfrak{a}C_\varpi\|w\|_2^2n^{-1}\ln(n),\end{aligned}\tag{3.26}$$

where we have denoted $C_\varpi=\int_{\mathbb{R}^{d-1}}\big(1+|\mathrm{t}+\mathrm{v}|_2^{\varpi}\big)^{-1}\mathrm{dv},\ \mathrm{x}=(x_2,\dots,x_d)^\top$, and $\mathrm{v}=(v_2,\dots,v_d)^\top$.

We deduce from (3.4.2) and (3.26) that for n sufficiently large

$$\sup_{j=1,\dots,N_n}\ \mathbb{E}_{F_0}^n\big\{Z^2\big(F_j,F_0\big)\big\}\leq n^{\mathfrak{a}\mathfrak{Q}\mathfrak{g}C_\varpi\|w\|_2^2}. \tag{3.27}$$

For any $j\neq k$ set $\Theta_{j,k}^\top=\big(\vartheta_j,\vartheta_k,\mathbf{e}_3,\dots,\mathbf{e}_d\big)$. By changing of variables with $\Theta_{j,k}(x-t)=v$ we have

$$\begin{aligned}&\int_{\mathbb{R}^d}\big|F_j(x)F_k(x)\big|\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x\\ \leq\ &|\det(\Theta_{j,k})|^{-1}L^2h^{2\beta}\int_{\mathbf{R}^d}\big|w\big(v_1/h\big)w\big(v_2/h\big)\big|\Big(1+\big|x+\Theta_{j,k}^{-1}v\big|_2^{\varpi}\Big)^{-1}\mathrm{d}v\\ \leq\ &|\det(\Theta_{j,k})|^{-1}c_\varpi L^2h^{2\beta+2}\|w\|_1^2,\end{aligned}$$

where we have put $c_\varpi=\int_{\mathbb{R}^{d-2}}\big(1+|\mathrm{x}+\mathrm{v}|_2^{\varpi}\big)^{-1}\mathrm{dv},\ \mathrm{v}=(v_3,\dots,v_d)^\top$ and $\mathrm{x}=(x_3,\dots,x_d)^\top$. Note that

$$\begin{aligned}|\det(\Theta_{j,k})| &= \big|\cos(j/N_n)\sin(k/N_n)-\cos(k/N_n)\sin(j/N_n)\big|\\ &= \big|\sin\big((k-j)/N_n\big)\big|\geq\sin\big(1/N_n\big)>(2N_n)^{-1},\end{aligned}$$

for sufficiently large n. Thus we finally get

$$\int_{\mathbb{R}^d} \big|F_j(x)F_k(x)\big|\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x \le 2\mathfrak{a}c_{\varpi}\|w\|_1^2 n^{-1}\ln(n)[N_n h].$$

Hence, choosing $b<2/(2\beta+1)$ we obtain for all n large enough

$$\sup_{j\neq k;\, j,k=1,\ldots,N_n} \int_{\mathbb{R}^d} \big|F_j(x)F_k(x)\big|\big(1+|x|_2^{\varpi}\big)^{-1}\mathrm{d}x \le 2\mathfrak{a}c_{\varpi}\|w\|_1^2 n^{-1}. \tag{3.28}$$

We deduce from (3.4.2) and (3.28)

$$\sup_{j\neq k;\, j,k=1,\ldots,N_n} \mathbb{E}^{(n)}_{F_0}\Big\{Z\big(F_j,F_0\big)Z\big(F_k,F_0\big)\Big\} \quad\le\quad e^{2\mathfrak{a}\mathfrak{Q}\mathfrak{g}c_{\varpi}\|w\|_1^2}. \tag{3.29}$$

We obtain from (3.27) and (3.29)

$$\mathbb{E}^{(n)}_{F_0}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} Z\big(F_j,F_0\big)\right)^2 \le n^{-b+\mathfrak{a}\mathfrak{Q}\mathfrak{g}C_{\varpi}\|w\|_2^2} + e^{2\mathfrak{a}\mathfrak{Q}\mathfrak{g}c_{\varpi}\|w\|_1^2}.$$

Choosing $\mathfrak{a} = b\big(\mathfrak{Q}\mathfrak{g}C_{\varpi}\|w\|_2^2\big)^{-1}$ we see that (3.21) holds with $c=1+e^{2\mathfrak{a}\mathfrak{Q}\mathfrak{g}c_{\varpi}\|w\|_1^2}$. Since that the constant c appeared in (3.21) is chosen independently of L, the assertion of the theorem follows from Lemma 4. ∎

3.4.3. *Proof of Theorem 5.* To prove the theorem we will exploit the ideas developed in Lepski et al. (1997). Moreover, our considerations are, to a great degree, based on the technical result of Lemma 5 below. Its proof is postponed until Appendix.

LEMMA 5. *Grant Assumptions 4 and 5. Then, for any $\mathfrak{p}>1$, $0<s\le \boldsymbol{b}$ and $\mathcal{Q}>0$ we have*

$$\sup_{g\in\mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})} \big\|\Delta^*_{\mathcal{K},g}(h,\cdot)\big\|_{\mathfrak{p}} \le 2\tau_{\mathfrak{p}}\mathcal{Q}h^s\|\mathcal{K}\|_\infty\left[2^{s\mathfrak{p}}-1\right]^{-\frac{1}{\mathfrak{p}}}, \quad \forall h>0.$$

Here $\tau_{\mathfrak{p}}$ is a depending only of $\mathfrak{p}$ constant from the $(\mathfrak{p},\mathfrak{p})$-strong maximal inequality.

Proof of Theorem 5. It is suffice to prove the theorem only in the case $r\ge p$. Indeed, remind that the risk $\mathcal{R}^{(n)}_r(\cdot,\cdot)$ is described by the L_r norm on $[-1/2,1/2]$, therefore

$$\mathcal{R}^{(n)}_r(\cdot,\cdot)\le \mathcal{R}^{(n)}_p(\cdot,\cdot), \quad r\le p.$$

Hence the case $r \le p$ can be reduced to the case $r = p$.

In view of Theorem 2 in order to obtain the assertion of the theorem it suffices to bound from above $\left\| \frac{\ln(n)}{nh^*_{\mathcal{K},f}(\cdot)} \right\|^{1/2}_{r/2}$.

Set $\Gamma_0 = \{y \in [-1/2, 1/2] : \quad h^*_{\mathcal{K},f}(y) = 1\}$ and $\Gamma_k = \{y \in [-1/2, 1/2] : h^*_{\mathcal{K},f}(y) \in (2^{-k}, 2^{-k+1}] \cap [h_{\min}, 1]\}$ for $k = 1, 2, \dots$. Later on, the integration over empty set is supposed to be zero. We have

$$\left\| \frac{\ln(n)}{nh^*_{\mathcal{K},f}(\cdot)} \right\|^{\frac{1}{2}}_{\frac{r}{2}} = \sum_{k \ge 1} \int_{\Gamma_k} \left(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(y)} \right)^{\frac{r}{2}} \mathrm{d}y + \int_{\Gamma_0} \left(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(y)} \right)^{\frac{r}{2}} \mathrm{d}y.$$

Later on $\overline{c}_i, i = 1, \dots,$ denote constants independent on n, f and L.

The definition of Γ_0 implies

$$\int_{\Gamma_0} \left(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(y)} \right)^{\frac{r}{2}} \mathrm{d}y \le \overline{c}_1 \left[n^{-1} \ln(n) \right]^{\frac{r}{2}}. \tag{3.30}$$

We have in view of (2.2) for any $k \ge 1$

$$\Delta^*_{\mathcal{K},f}\big(h^*_{\mathcal{K},f}(y), y\big) = \left[\frac{D^2(\omega) \|\mathcal{K}\|^2_\infty \ln(n)}{nh^*_{\mathcal{K},f}(y)} \right]^{\frac{1}{2}}, \quad \forall y \in \Gamma_k. \tag{3.31}$$

Let $0 \le q_k \le r$ be a sequence whose choice will be done later. We obtain from (3.31)

$$\begin{aligned} \sum_{k \ge 1} \int_{\Gamma_k} \left(\frac{\ln(n)}{nh^*_{\mathcal{K},f}(y)} \right)^{\frac{r}{2}} \mathrm{d}y &\le \overline{c}_2 \sum_{k \ge 1} \left(\frac{\ln(n)}{n2^{-k}} \right)^{\frac{r-q_k}{2}} \int_{\Gamma_k} \left(\Delta^*_{\mathcal{K},f}\big(2^{1-k}, y\big) \right)^{q_k} \mathrm{d}y \\ &\le \overline{c}_2 \sum_{k \ge 1} \left(\frac{\ln(n)}{n2^{-k}} \right)^{\frac{r-q_k}{2}} \int \left(\Delta^*_{\mathcal{K},f}\big(2^{1-k}, y\big) \right)^{q_k} \mathrm{d}y =: \Xi. \end{aligned} \tag{3.32}$$

To get the first inequality we have used that $\Delta^*_{\mathcal{K},f}\big(\cdot, y\big)$ in monotonically increasing function.

The computation of the quantity on the right-hand side of (3.32), including the choice of $(q_k,\ k \ge 1)$, will be done differently in dependence on β, p and r.

1^0. *Case* $(2\beta + 1)p > r$. Put $h^* = \left[L^{-2} n^{-1} \ln(n) \right]^{\frac{1}{2\beta+1}}$ and choose $q_k = p$ if $2^{-k} \le h^*$ and $q_k = 0$ if $2^{-k} > h^*$.

Applying Lemma 5 with $\mathfrak{p}=p$, $s=\beta$ and $\mathcal{Q}=L$ we get

$$\begin{aligned}\Xi &\le \overline{c}_3 L^p \sum_{k:\,2^{-k}\le h^*}\left(\frac{\ln(n)}{n2^{-k}}\right)^{\frac{r-p}{2}} 2^{-k\beta p}+\overline{c}_4\left(\frac{\ln(n)}{nh^*}\right)^{\frac{r}{2}}\\ &\le \overline{c}_5\left[L^p\left(n^{-1}\ln(n)\right)^{\frac{r-p}{2}}\sum_{k:\,2^{-k}\le h^*}2^{-k\left[\beta p-\frac{r-p}{2}\right]}+\left(\frac{\ln(n)}{nh^*}\right)^{\frac{r}{2}}\right]. \qquad (3.33)\end{aligned}$$

Because in the considered case $\beta p-\frac{r-p}{2}>0$, we obtain

$$\Xi\le\overline{c}_6\left[L^p\left(\ln(n)\right)^{\frac{r-p}{2}}(h^*)^{\beta p-\frac{r-p}{2}}+\left(\frac{\ln(n)}{nh^*}\right)^{\frac{r}{2}}\right].$$

It remains to note that h^* is chosen by balancing two terms on the right-hand side of the latter inequality. It yields

$$\Xi\le\overline{c}_7 L^{\frac{r}{2\beta+1}}\left(n^{-1}\ln(n)\right)^{\frac{r\beta}{2\beta+1}}. \qquad (3.34)$$

The argument in the case $(2\beta+1)p>r$ is completed with the use of Theorem 2, (3.30) and (3.34).

2^0. *Case* $(2\beta+1)p=r$. Put $h^*=1$ and choose $q_k=p$ for all $k\ge1$. Repeating the computations led to (3.33) we get

$$\Xi\le\overline{c}_8\ln(n)L^p\left(n^{-1}\ln(n)\right)^{\frac{r-p}{2}}. \qquad (3.35)$$

Here we have used that $\beta p-\frac{r-p}{2}=0$ and that the summation in (3.32) runs over k such that $2^{-k}\ge h_{\min}$, since otherwise $\Gamma_k=\emptyset$. It remains to note that the equality $(2\beta+1)p=r$ is equivalent to $p/r=1/(2\beta+1)$ and $(r-p)/2r=\beta/(2\beta+1)$. The assertion of the theorem in the case $(2\beta+1)p=r$ follows now from Theorem 2, (3.30) and (3.35).

3^0. *Case* $(2\beta+1)p<r$. Choose $q_k=r$ if $2^{-k}\le h^*$ and $q_k=p$ if $2^{-k}>h^*$, where the choice of h^* will be done later.

The following embedding holds, see Besov et al. (1979): $\mathbb{N}_p(\beta,L)\subseteq\mathbb{N}_r\big(\beta-1/p+1/r,c_6L\big)$. Thus, applying Lemma 5 with $\mathfrak{p}=r$, $s=\beta-1/p+1/r$ and $\mathcal{Q}=c_6L$ we get

$$\begin{aligned}\Xi_1 &:= \sum_{k:\,2^{-k}\le h^*}\left(\frac{\ln(n)}{n2^{-k}}\right)^{\frac{r-q_k}{2}}\int\Big(\Delta^*_{\mathcal{K},f}\big(2^{1-k},y\big)\Big)^{q_k}\mathrm{d}y\\ &= \sum_{k:\,2^{-k}\le h^*}\int\Big(\Delta^*_{\mathcal{K},f}\big(2^{1-k},y\big)\Big)^{r}\mathrm{d}y\le\overline{c}_9L^r(h^*)^{\beta r-(r/p)+1}. \qquad (3.36)\end{aligned}$$

Applying Lemma 5 with $\mathfrak{p} = r$, $s = \beta$ and $\mathcal{Q} = L$ we get

$$\begin{aligned}\Xi_2 \quad &:= \sum_{k:\, 2^{-k} > h^*} \left(\frac{\ln(n)}{n2^{-k}}\right)^{\frac{r-q_k}{2}} \int \left(\Delta^*_{\mathcal{K},f}(2^{1-k}, y)\right)^{q_k} \mathrm{d}y \\ &= \bar{c}_{10} L^p \left(n^{-1}\ln(n)\right)^{\frac{r-p}{2}} \sum_{k:\, 2^{-k} > h^*} 2^{-k\left[\beta p - \frac{r-p}{2}\right]} \\ &\leq \bar{c}_{11} L^p \left(n^{-1}\ln(n)\right)^{\frac{r-p}{2}} (h^*)^{\beta p - \frac{r-p}{2}}. \end{aligned} \tag{3.37}$$

Here we have used that $\beta p - \frac{r-p}{2} < 0$. In view of (3.36) and (3.37) we choose h^* from the equality:

$$L^r (h^*)^{\beta r - (r/p) + 1} = L^p \left(n^{-1}\ln(n)\right)^{\frac{r-p}{2}} (h^*)^{\beta p - \frac{r-p}{2}}.$$

It yields $h^* = \left(L^{-1} n^{-1} \ln(n)\right)^{\frac{1}{2\beta - 2/p + 1}}$ and we obtain finally that

$$\Xi \leq \bar{c}_{12} L^{\frac{r(1/2 - 1/r)}{\beta - 1/p + 1/2}} \left(n^{-1}\ln(n)\right)^{\frac{r(\beta - 1/p + 1/r)}{2\beta - 2/p + 1}}. \tag{3.38}$$

The assertion of the theorem in the case $(2\beta + 1)p < r$ follows now from Theorem 2, (3.30) and (3.38). ∎

4. Appendix.

4.1. *Proof of Lemma 1.*

Proof of the first assertion. The symmetry of the kernel $\mathcal{K}$, see Assumption 4, implies

$$S_{(-\theta^*,h)(\nu,h)}(\cdot) \equiv S_{(\theta^*,h)(-\nu,h)}(\cdot), \quad S_{(-\nu,h)}(\cdot) \equiv S_{(\nu,h)}(\cdot).$$

Therefore it suffices to prove the first assertion of the lemma under the condition $\nu^\top \theta^* \geq 0$. In this case $E_{(\theta^*,h)(\nu,h)} = \overline{E}_{(\theta^*,h)(\nu,h)}$ and we note that

$$\overline{E}_{(\theta^*,h)(\nu,h)} = \left[E^{-1}_{(\theta^*,h)} + E^{-1}_{(\nu,h)}\right]^{-1}. \tag{4.1}$$

For any $\theta = (\theta_1, \theta_2) \in \mathbb{S}^1$ let $\theta_\perp = (-\theta_2, \theta_1)$. Using (4.1) we obtain

$$\begin{aligned} S_{(\theta^*,h)(\nu,h)}(x) &= \int K(u) f\left(h[\theta^* + \nu]^\top \theta^* u_1 + [\theta^*{}_\perp + \nu_\perp]^\top \theta^* u_2 + x^\top \theta^*\right) \mathrm{d}u \\ &= \int\int \mathcal{K}(u_1)\mathcal{K}(u_2) f\left(h[1 + \nu^\top \theta^*] u_1 + \nu_\perp^\top \theta^* u_2 + x^\top \theta^*\right) \mathrm{d}u_1 \mathrm{d}u_2. \end{aligned}$$

We also have

$$S_{(\nu,h)}(x) = \int\int \mathcal{K}(u_1)\mathcal{K}(u_2) f\big(h\nu^\top\theta^* u_1 + \nu_\perp^\top\theta^* u_2 + x^\top\theta^*\big)\mathrm{d}u_1\mathrm{d}u_2.$$

Put $S^*_\nu(x) = \int \mathcal{K}(u_2) f\big(\nu_\perp^\top\theta^* u_2 + x^\top\theta^*\big)\mathrm{d}u_2$ and consider two cases.

1^0. $\nu_\perp^\top\theta^* = 0$. In this case $S^*_\nu(x) = f(x^\top\theta^*)$ and

$$S_{(\nu,h)}(x) = \int \mathcal{K}(u_1) f\big(hu_1 + x^\top\theta^*\big)\mathrm{d}u_1 = \frac{1}{h}\int \mathcal{K}\bigg(\frac{t - x^\top\theta^*}{h}\bigg) f(t)\mathrm{d}t,$$

$$S_{(\theta^*,h)(\nu,h)}(x) = \int \mathcal{K}(u_1) f\big(2hu_1 + x^\top\theta^*\big)\mathrm{d}u_1 = \frac{1}{2h}\int \mathcal{K}\bigg(\frac{t - x^\top\theta^*}{2h}\bigg) f(t)\mathrm{d}t.$$

Here we have used that $\nu_\perp^\top\theta^* = 0$ together with $\nu^\top\theta^* \geq 0$ implies $\nu = \theta^*$. Thus, we obtain

$$\begin{aligned}
&\big|S_{(\theta^*,h)(\nu,h)}(x) - S_{(\nu,h)}(x)\big| \\
&\leq \big|S_{(\theta^*,h)(\nu,h)}(x) - S^*_\nu(x)\big| + \big|S_{(\nu,h)}(x) - S^*_\nu(x)\big| \\
&\leq \Delta_{\mathcal{K},f}\big(h, x^\top\theta^*\big) + \Delta_{\mathcal{K},f}\big(2h, x^\top\theta^*\big) \leq 2\Delta^*_{\mathcal{K},f}\big(2h, x^\top\theta^*\big).
\end{aligned}\tag{4.2}$$

2^0. $\nu_\perp^\top\theta^* \neq 0$. In this case we have

$$S^*_\nu(x) = \int\int \frac{\mathcal{K}\Big(\frac{v_1}{h(1+\nu^\top\theta^*)}\Big)\mathcal{K}\Big(\frac{v_2 - x^\top\theta^*}{|\nu_\perp^\top\theta^*|}\Big)}{h(1+\nu^\top\theta^*)|\nu_\perp^\top\theta^*|} f(v_2)\mathrm{d}v_1\mathrm{d}v_2,$$

$$S_{(\vartheta^*,h)(\nu,h)}(x) = \int\int \frac{\mathcal{K}\Big(\frac{v_1}{h(1+\nu^\top\theta^*)}\Big)\mathcal{K}\Big(\frac{v_2 - x^\top\theta^*}{|\nu_\perp^\top\theta^*|}\Big)}{h(1+\nu^\top\theta^*)|\nu_\perp^\top\theta^*|} f(v_1+v_2)\mathrm{d}v_1\mathrm{d}v_2.$$

Here we have used once again the symmetry of $\mathcal{K}$. Thus, taking into account that $|\nu^\top\theta^*| \leq 1$, we get

$$\begin{aligned}
&\big|S_{(\theta^*,h)(\nu,h)}(x) - S^*_\nu(x)\big| \\
&\leq \int \frac{1}{|\nu_\perp^\top\theta^*|}\left|\mathcal{K}\left(\frac{v_2 - x^\top\theta^*}{|\nu_\perp^\top\theta^*|}\right)\right| \sup_{\delta\leq 2h}\left|\int \frac{1}{\delta}\mathcal{K}\left(\frac{v_1}{\delta}\right)\big[f(v_1+v_2) - f(v_2)\big]\mathrm{d}v_1\right|\mathrm{d}v_2 \\
&\leq \|\mathcal{K}\|_\infty \sup_{a>0}\left[\frac{1}{a}\int_{x^\top\theta^* - a/2}^{x^\top\theta^* + a/2} \sup_{\delta\leq 2h}\left|\int \frac{1}{\delta}\mathcal{K}\left(\frac{v_1}{\delta}\right)\big[f(v_1+v_2) - f(v_2)\big]\mathrm{d}v_1\right|\mathrm{d}v_2\right].
\end{aligned}$$

Here we have used that $\mathrm{supp}(\mathcal{K}) \subseteq [-1/2, 1/2]$ (Assumption 4 (*1*)). Hence,

$$\big|S_{(\theta^*,h)(\nu,h)}(x) - S^*_\nu(x)\big| \leq \|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(2h, x^\top\theta^*\big).\tag{4.3}$$

If $\nu^\top\theta^* \neq 0$ we obtain by the same computations

$$\big|S_{(\nu,h)}(x) - S^*_\nu(x)\big| \leq \|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(h, x^\top\theta^*\big).$$

Noting that $S_{(\nu,h)}(\cdot) \equiv S^*_\nu(\cdot)$ if $\nu^\top\theta^* = 0$ we get

$$\big|S_{(\nu,h)}(x) - S^*_\nu(x)\big| \leq \|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(h, x^\top\theta^*\big), \tag{4.4}$$

that yields together with (4.3)

$$\big|S_{(\theta^*,h)(\nu,h)}(x) - S_{(\nu,h)}(x)\big| \leq 2\|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(2h, x^\top\theta^*\big). \tag{4.5}$$

Finally, taking into account that in view of Assumption 4 (*1*) $\|\mathcal{K}\|_\infty \geq 1$, we obtain from (4.2) and (4.5) that

$$\begin{aligned}\big|S_{(\theta^*,h)(\nu,h)}(x) - S_{(\nu,h)}(x)\big| &\leq 2\|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(2h, x^\top\theta^*\big)\\ &\leq 2\|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(h^*_f, x^\top\theta^*\big),\end{aligned}$$

since we consider h such that $2h \leq h^*_f$. The definition of h^*_f implies

$$\Delta^*_{\mathcal{K},f}\big(h^*_f, x^\top\theta^*\big) \leq (h^*_f)^{-1/2}\|\mathcal{K}\|_\infty\sqrt{n^{-1}\ln(n)}$$

and the first assertion of the lemma follows.

Proof of the second and third assertions. In view of (4.4) for $\forall\eta \leq h \leq h^*_f$

$$\begin{aligned}&\big|S_{(\nu,\eta)}(x) - S_{(\nu,h)}(x)\big| \leq \big|S_{(\nu,\eta)}(x) - S^*_\nu(x)\big| + \big|S_{(\nu,h)}(x) - S^*_\nu(x)\big|\\ &\leq \|\mathcal{K}\|_\infty\Big[\Delta^*_{\mathcal{K},f}\big(\eta, x^\top\theta^*\big) + \Delta^*_{\mathcal{K},f}\big(h, x^\top\theta^*\big)\Big] \leq 2\|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(h, x^\top\theta^*\big)\\ &\leq 2\|\mathcal{K}\|_\infty \Delta^*_{\mathcal{K},f}\big(h^*_f, x^\top\theta^*\big) \leq 2(h^*_f)^{-1/2}\|\mathcal{K}\|^2_\infty\sqrt{n^{-1}\ln(n)},\end{aligned}$$

in view of the definition of h^*_f. The second assertion is proved.

We have for any $h \leq h^*_f$

$$\begin{aligned}\big|S_{(\theta^*,h)}(x) - F(x)\big| &= \left|\frac{1}{h}\int \mathcal{K}\left(\frac{u}{h}\right)\big[f(u + x^\top\theta^*) - f(x^\top\theta^*)\big]\mathrm{d}u\right|\\ &\leq \Delta_{\mathcal{K},f}\big(h, x^\top\theta^*\big) \leq \Delta^*_{\mathcal{K},f}\big(h, x^\top\theta^*\big) \leq \Delta^*_{\mathcal{K},f}\big(h^*_f, x^\top\theta^*\big)\\ &= (h^*_f)^{-1/2}\|\mathcal{K}\|_\infty\varepsilon\sqrt{n^{-1}\ln(n)},\end{aligned}$$

in view of the definition of h^*_f. The third assertion is proved. ∎

4.2. *Proof of Lemma 2.* The proof of the lemma is based on auxiliary result (4.12) justified in part 1^0 below.

1^0. Let $(\mathbb{V},\mathfrak{m})$ be a measurable space and let $\mathcal{Z}\subseteq[-B,B]^{\mathfrak{s}},\mathfrak{s}\geq 1$, for some given $B\geq 1$. Let $V_1,\dots,V_n$ be i.i.d. $\mathbb{V}$-valued random variables and later on P denotes the probability law of $V_1,\dots,V_n$.

Suppose we are given by $G:\mathbb{V}\times\mathcal{Z}\to\mathbb{R}$ and consider the random field

$$\zeta_n(z)=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big[G(V_i,z)-\mathrm{E}G(V_i,z)\big],\quad z\in\mathcal{Z}.$$

Below we will be interested in establishing of a tail probability inequality for $\sup_{z\in\mathcal{Z}}\big|\zeta_n(z)\big|$. To do that we will apply Proposition 1 in Lepski (2013) with the functional $\Psi(\cdot)=|\cdot|$. It is convenient to impose the following assumptions.

$$\sup_{v\in\mathbb{V}}\sup_{z\in\mathcal{Z}}|G(v,z)|=:G^*_\infty<\infty. \tag{4.6}$$

There exist $\alpha\in(0,1]$ and $R\geq 1$ such that

$$b_\infty(z,z'):=\sup_{v\in\mathbb{V}}\big|G(v,z)-G(v,z')\big|\leq R|z-z'|^\alpha_\infty,\quad \forall z,z'\in\mathcal{Z}. \tag{4.7}$$

Here $|\cdot|_\infty$ denotes the vector supremum norm on $\mathbb{R}^{\mathfrak{s}}$. Note that

$$\mathrm{a}(z,z'):=\sqrt{2\mathrm{E}|G(V_1,z)-G(V_1,z')|^2}\leq\sqrt{2}b_\infty(z,z'). \tag{4.8}$$

Set $\mathrm{b}=(4/3)n^{-\frac{1}{2}}b_\infty$ and equip $\mathcal{Z}$ with the semi-norms a and b. Denote also by $\mathfrak{E}_\mathrm{a}(\epsilon),\ \mathfrak{E}_\mathrm{b}(\epsilon),\epsilon>0$, the ϵ-entropy of $\mathcal{Z}$ measured in the metrics generated by a and b respectively. Put finally $\mathfrak{E}(\delta)$, the δ-entropy of $[-B,B]^{\mathfrak{s}}$ measured in $|\cdot|_\infty$. Since obviously $\mathfrak{E}(\delta)\leq\mathfrak{s}\big[\ln\big(B/\delta\big)\big]_+$, we obtain in view of (4.7) and (4.8) for any $\epsilon>0$, taking into account that $B\geq 1,\alpha\leq 1$,

$$\mathfrak{E}_\mathrm{a}(\epsilon)\;\leq\;\mathfrak{E}\Big([2\sqrt{2}R]^{-1/\alpha}\epsilon^{1/\alpha}\Big)\leq\frac{\mathfrak{s}}{\alpha}\Big[\ln\big(2\sqrt{2}BR\epsilon^{-1}\big)\Big]_+; \tag{4.9}$$

$$\mathfrak{E}_\mathrm{b}(\epsilon)\;\leq\;\mathfrak{E}\big([8R/3]^{-1/\alpha}[\sqrt{n}\epsilon]^{1/\alpha}\big)\leq\frac{\mathfrak{s}}{\alpha}\Big[\ln\big([8/3]BRn^{-\frac{1}{2}}\epsilon^{-1}\big)\Big]_+, \tag{4.10}$$

where $[t]_+$ denotes the positive part of t. Here we have used also the fact that ϵ-entropy of $\mathcal{Z}$ is less than $(\epsilon/2)$-entropy of $[-B,B]^{\mathfrak{s}}$ whatever is the metric in which these entropies are measured.

Set $s(u)=(3/4)u^{-4},u>0$, and introduce the quantities

$$e_\mathrm{a}(x)=\sup_{\delta>0}\delta^{-2}\mathfrak{E}_\mathrm{a}\Big(\frac{xs(\delta)}{48\delta}\Big),\qquad e_\mathrm{b}(x)=\sup_{\delta>0}\delta^{-1}\mathfrak{E}_\mathrm{b}\Big(\frac{xs(\delta)}{48\delta}\Big),\quad x>0.$$

First, we note that the bounds (4.9) and (4.10) guarantee, for any $x>0$,

$$e_{\mathrm{a}}(x)<\infty, \qquad e_{\mathrm{b}}(x)<\infty. \tag{4.11}$$

Next, denoting $G_2^*=\sup_{z\in\mathcal{Z}}\big(\mathrm{E}|G(V_1,z)|^2\big)^{1/2}$, we get in view of the triangle inequality

$$\sup_{z,z'\in\mathcal{Z}}\mathrm{a}(z,z')\leq 2\sqrt{2}G_2^*, \qquad \sup_{z,z'\in\mathcal{Z}}\mathrm{b}(z,z')\leq (8/3)n^{-\frac{1}{2}}G_\infty^*,$$

that implies that $\mathfrak{E}_{\mathrm{a}}(\epsilon)=0$ for all $\epsilon\geq 2\sqrt{2}G_2^*$, and $\mathfrak{E}_{\mathrm{b}}(\epsilon)=0$, for all $\epsilon\geq(8/3)n^{-\frac{1}{2}}G_\infty^*$. Hence, for all $\varkappa_1\geq G_2^*$ and $\varkappa_2\geq G_\infty^*$, we obtain

$$\begin{aligned} e_{\mathrm{a}}\Big(2\sqrt{2}\varkappa_1\Big) &= \sup_{\delta\geq\delta_0}\delta^{-2}\mathfrak{E}_{\mathrm{a}}\Big(2\sqrt{2}\varkappa_1(48\delta)^{-1}s(\delta)\Big);\\ e_{\mathrm{b}}\Big((8/3)n^{-\frac{1}{2}}\varkappa_2\Big) &= \sup_{\delta\geq\delta_0}\delta^{-1}\mathfrak{E}_{\mathrm{b}}\Big((8/3)n^{-\frac{1}{2}}\varkappa_2(48\delta)^{-1}s(\delta)\Big),\end{aligned}$$

where $\delta_0=2^{-\frac{6}{5}}$. That together with (4.9) and (4.10) leads to

$$\begin{aligned} e_{\mathrm{a}}\Big(2\sqrt{2}\varkappa_1\Big) &\leq (\mathfrak{s}/\alpha)\Big\{5.3\big[\ln\big(BR\varkappa_1^{-1}\big)\big]_+ + 3.2\Big\};\\ e_{\mathrm{b}}\Big((8/3)n^{-\frac{1}{2}}\varkappa_2\Big) &\leq (\mathfrak{s}/\alpha)\Big\{2.3\big[\ln\big(BR\varkappa_2^{-1}\big)\big]_+ + 0.85\Big\}.\end{aligned}$$

Put finally $e(\varkappa_1,\varkappa_2)=(\mathfrak{s}/\alpha)\Big\{7.6\big[\ln\big(BR[\varkappa_1\wedge\varkappa_2]^{-1}\big)\big]_+ + 4.1\Big\}$ and define for any $r\geq 1$

$$\begin{aligned} U_n(\varkappa_1,\varkappa_2) &= 2\sqrt{2}\varkappa_1\sqrt{24e(\varkappa_1,\varkappa_2)+8\,r\ln(n)}\\ &\quad+\frac{8}{3\sqrt{n}}\varkappa_2\Big[24e(\varkappa_1,\varkappa_2)+8\,r\ln(n)\Big].\end{aligned}$$

It remains to note that the verification of the assumptions of Proposition 1 in Lepski (2013) follows from (4.6), (4.8), (4.11) and Bernstein's inequality. Applying the proposition with $\Psi(\cdot)=|\cdot|$, $\varepsilon=\sqrt{2}-1$ and $y=8\,r\ln(n)$ we obtain for any $n\geq 3$ and any $r\geq 1$

$$\mathrm{P}\bigg\{\sup_{z\in\mathcal{Z}}|\zeta_n(z)|\geq U_n(\varkappa_1,\varkappa_2)\bigg\}\leq 4n^{-4r}. \tag{4.12}$$

2^0. Let us prove that

$$\mathbb{P}_X^{(n)}\bigg\{\sup_{E\in\mathcal{E}_*}\big|\eta_{n,t}(E)\big|\geq C_1(n)\|F\|_\infty\bigg\}\leq 4n^{-4r}. \tag{4.13}$$

Considering a 2×2-matrix as an element of $\mathbb{R}^4$ we obviously have that the defined in (3.2) set of matrices $\mathcal{E}_{a,A}\subset[-B,B]^4$, with $B=(\sqrt{2a})^{-1}A$. We apply the result from 1^0 with $\mathbb{V}=\mathbb{R}^2$, $\mathcal{Z}=\mathcal{E}_{a,A}$, $z=(e_{11},e_{12},e_{21},e_{22})$, where e_{ij} are the entries of the matrix E, $\mathfrak{s}=4$, and $G(\cdot,z)=J(\cdot,E)F(\cdot)$ that corresponds to $\{\zeta_n(z),z\in\mathcal{Z}\}=\{\eta_{n,t}(E),E\in\mathcal{E}_{a,A}\}$.

First, we note that

$$\big|\det(E)\big|\geq a,\quad \forall E\in\mathcal{E}_{a,A}. \tag{4.14}$$

Indeed, $\big|\det(E)\big|\leq 2|E|^2_\infty\leq 2(\sqrt{2a})^{-2}\big|\det(E)\big|^2=a^{-1}\big|\det(E)\big|^2$, and (4.14) follows.

Next, in the considered case, since $\mathrm{supp}(\mathcal{K})\subseteq[-1/2,1/2]^2$, after changing the variables we have

$$(G^*_2)^2=\sup_{E\in\mathcal{E}_{a,A}}\int_{[-1/2,1/2]^2}K^2(u)F^2\big(t+E^{-1}u\big)g^{-1}\big(t+E^{-1}u\big)\mathrm{d}u.$$

Note that $\big|E^{-1}\big|_\infty=\big|\det(E)\big|^{-1}\big|E\big|_\infty\leq(\sqrt{2a})^{-1}$ for any $E\in\mathcal{E}_{a,A}$ and, therefore,

$$\sup_{u\in[-1/2,1/2]^2}F^2\big(t+E^{-1}u\big)g^{-1}\big(t+E^{-1}u\big)\leq\sup_{x\in\Delta_a}F^2(x)g^{-1}(x)=:\mathcal{R}^2_{2,\mathrm{a}}(F,g),$$

where $\Delta_a=\big[-\frac12-\frac{1}{\sqrt{2a}},\frac12+\frac{1}{\sqrt{2a}}\big]^2$. Thus,

$$G^*_2\leq\|\mathcal{K}\|^2_2\mathcal{R}_{\mathrm{a},2}(F,g). \tag{4.15}$$

By the same reasons denoting $\mathcal{R}_{\mathrm{a},\infty}(F,g)=\sup_{x\in\Delta_a}|F(x)|g^{-1}(x)$ we obtain

$$G^*_\infty\leq\sqrt{A}\|\mathcal{K}\|^2_\infty\mathcal{R}_{\mathrm{a},\infty}(F,g). \tag{4.16}$$

Since $F\in\mathbb{F}(\beta_0,M)$ assumption (4.6) is fulfilled as soon as $\inf_{x\in\Delta_a}g(x)<\infty$.

It remains to check (4.7). By the use of the triangle inequality we get

$$\begin{aligned}
b_\infty\big(E,E'\big) &:= \sup_{x\in\mathbb{R}^2}\big|J(x,E)-J(x,E')\big|\,|F(x)|\\
&\leq \Big|\sqrt{\big|\det(E)\big|}-\sqrt{\big|\det(E')\big|}\Big|\|\mathcal{K}\|^2_\infty\mathcal{R}_{\mathrm{a},\infty}(F,g)\\
&+ \sqrt{A}\sup_{x\in\mathbb{R}^2}\Big\{\big|K\big(E(x-t)-K\big(E'(x-t)\big)\big||F(x)|g^{-1}(x)\Big\}
\end{aligned} \tag{4.17}$$

First, we note that for any $E,E'\in\mathcal{E}_{a,A}$

$$K\big(E(x-t)=0,\quad K\big(E'(x-t)=0,\ \ \forall x\neq\Delta_a.$$

Second, Assumption 4 yields $\big|K(y)-K(z)\big| \le 2Q\|\mathcal{K}\|_\infty |y-z|_\infty$, $\forall y,z\in\mathbb{R}^2$, that together with the previous display gives

$$\begin{aligned}
&\sqrt{A}\sup_{x\in\mathbb{R}}\Big\{\big|K\big(E(x-t)-K\big(E'(x-t)\big)\big||F(x)|g^{-1}(x)\Big\}\\
&\le 4\sqrt{A}Q\|\mathcal{K}\|_\infty\big[1+(\sqrt{2a})^{-1}\big]\mathcal{R}_{\mathrm{a},\infty}(F,g)\big|E-E'\big|_\infty\\
&\le 12Q\|\mathcal{K}\|^2_\infty(A/a)\mathcal{R}_{\mathrm{a},\infty}(F,g)\big|E-E'\big|_\infty. \qquad (4.18)
\end{aligned}$$

Here we have also used that $a\le 1$, $A\ge 1$ and $\|\mathcal{K}\|_\infty\ge 1$. Next, using obvious for 2×2-matrices inequality $|\det(E)-\det(E')|\le 4\,[|E|_\infty\vee|E'|_\infty]\,\big|E-E'\big|_\infty$ we get in view of (4.14)

$$\bigg|\sqrt{\big|\det(E)\big|}-\sqrt{\big|\det(E')\big|}\bigg|\le\sqrt{2}(A/a)\big|E-E'\big|_\infty,\quad\forall E,E'\in\mathcal{E}_{a,A}. \qquad (4.19)$$

We conclude in view of (4.17), (4.18) and (4.19) that (4.7) is fulfilled with

$$\alpha=1,\qquad R=(A/a)\mathcal{R}_{\mathrm{a},\infty}(F,g)\|\mathcal{K}\|^2_\infty\big[12Q+\sqrt{2}\big]. \qquad (4.20)$$

Let now $a=1/8$, $A=(h_{\min})^{-1}=n\big(\ln(n)\big)^{-\frac{2+\omega}{\omega}}$.

First we note that $\Delta_a=\big[-5/2,5/2\big]^2$ and, therefore, in view of Assumption 2 we have

$$\mathcal{R}_{\mathrm{a},2}(F,g)\le\underline{g}^{-\frac12}||F||_\infty\quad\mathcal{R}_{\mathrm{a},\infty}(F,g)\le\underline{g}^{-1}||F||_\infty.$$

It yields together with (4.15) and (4.16)

$$\begin{aligned}
&G^*_2\le\underline{g}^{-\frac12}\|\mathcal{K}\|^2_\infty||F||_\infty=:\varkappa_1,\\
&G^*_\infty\le n^{\frac12}\big(\ln(n)\big)^{-\frac{2+\omega}{2\omega}}\underline{g}^{-1}\|\mathcal{K}\|^2_\infty||F||_\infty=:\varkappa_2. \qquad (4.21)
\end{aligned}$$

Here we have also used that $\|\mathcal{K}\|_2\le\|\mathcal{K}\|_\infty$ in view of Assumption 4. Taking into account that $h_{\min}>n^{-1}$ we obtain from (4.20)

$$R\le 8n\underline{g}^{-1}||F||_\infty\|\mathcal{K}\|^2_\infty\big[12Q+\sqrt{2}\big], \qquad (4.22)$$

Thus, putting

$$c_1(n)=730\ln\Big(16n^2\underline{g}^{-\frac12}\big[12Q+\sqrt{2}\big]\Big)+8r\ln(n)+394,$$

we deduce from (4.21) and (4.22) that (4.12) holds with

$$\begin{aligned}
U_n(\varkappa_1,\varkappa_2) &= 2\sqrt{2}\underline{g}^{-\frac12}\|\mathcal{K}\|^2_\infty||F||_\infty\sqrt{c_1(n)}\\
&\quad+(8/3)\big(\ln(n)\big)^{-\frac{2+\omega}{2\omega}}\underline{g}^{-1}\|\mathcal{K}\|^2_\infty||F||_\infty c_1(n).
\end{aligned}$$

It remains to note that $U_n(\varkappa_1, \varkappa_2) = ||F||_\infty C_1(n)$ and (4.13) follows.

3^0. Let us prove that

$$\mathbb{P}^{(n)}_{X,\varepsilon}\Big\{ \sup_{E\in\mathcal{E}_*} \big|\xi_{n,t}(E)\big| \geq C_2(n)\Big\} \leq (4+\Upsilon)n^{-4r}; \tag{4.23}$$

Let $v = (x,w) \in \mathbb{V} := \mathbb{R}^3$ and set for given $\tau > 0$

$$\begin{aligned} I_1(v,E) &= \sqrt{\big|\det(E)\big|}K\big(E(x-t)\big)g^{-1}(x)w1_{[-\tau,\tau]}(w); \\ I_2(v,E) &= \sqrt{\big|\det(E)\big|}K\big(E(x-t)\big)g^{-1}(x)w\Big[1 - 1_{[-\tau,\tau]}(w)\Big]. \end{aligned}$$

Putting $V_i = (X_i, \varepsilon_i), i = 1, \dots, n$, we remark that for any $\tau > 0$ and any $E \in \mathcal{E}_{a,A}$

$$\xi_{n,t}(E) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{I_1(V_i,E) + I_2(V_i,E)\} =: \xi^{(1)}_{n,t}(E) + \xi^{(2)}_{n,t}(E).$$

Noting that $\Big\{\xi^{(2)}_{n,t}(E) = 0\Big\} \supseteq \{\max_{i=1,\dots,n} |\varepsilon_i| < \tau\}$ we have for any $T > 0$ and any $\tau > 0$

$$\begin{aligned} &\mathbb{P}^{(n)}_{X,\varepsilon}\bigg(\sup_{E\in\mathcal{E}_{a,A}} \big|\xi_{n,t}(E)\big| \geq T\bigg) \\ \leq\; &\mathbb{P}^{(n)}_{X,\varepsilon}\bigg(\sup_{E\in\mathcal{E}_{a,A}} \big|\xi^{(1)}_{n,t}(E)\big| \geq T\bigg) + n\Upsilon e^{-\Omega\tau^\omega} \end{aligned} \tag{4.24}$$

in view of Assumption 1. Thus, it is sufficient to apply the result obtained in the step 1^0 to the random field $\xi^{(1)}_{n,t}(E)$. It is eligible because the independence of $\{\varepsilon_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ and symmetry of ε_1 provide that $\mathbb{E}^{(n)}_{X,\varepsilon}\big[\xi^{(1)}_{n,t}(E)\big] = 0$. Moreover, we deduce from (4.15) and (4.16)

$$G^*_2 \leq \sigma\|\mathcal{K}\|^2_2 \mathcal{R}_{\mathrm{a},2}(\mathbf{1},g), \qquad G^*_\infty \leq \tau\sqrt{A}\|\mathcal{K}\|^2_\infty \mathcal{R}_{\mathrm{a},\infty}(\mathbf{1},g),$$

where $\mathbf{1}$ denotes the function identically equal to 1. Remind also that $\sigma^2 = \sup_{p\in\mathfrak{P}} \int_{\mathbb{R}} x^2 p(x)\mathrm{d}x$. Note that for any $v = (x,w)$

$$\big|I_1(v,E) - I_1(v,E')\big| \leq \tau\,\big|J(x,E) - J(x,E')\big|$$

and, therefore, we obtain from (4.20) that (4.7) is fulfilled with

$$\alpha = 1, \qquad R = \tau(A/a)\mathcal{R}_{\mathrm{a},\infty}(\mathbf{1},g)\|\mathcal{K}\|^2_\infty\big[12Q + \sqrt{2}\big].$$

Let $a = 1/8, A = h_{\min}^{-1} = n\big(\ln(n)\big)^{-(2+\omega)/\omega}$ and choose the truncation level $\tau = \big(\Omega^{-1}(4r+1)\ln(n)\big)^{1/\omega}$. We have

$$\begin{aligned} & G_2^* \le (\sigma \vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2 =: \varkappa_1, \\ (4.25) \qquad & G_\infty^* \le n^{\frac{1}{2}}\big(\Omega^{-1}(4r+1)\big)^{\frac{1}{\omega}}\big(\ln(n)\big)^{-\frac{1}{2}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2 =: \varkappa_2. \end{aligned}$$

Here we have also used that $\|\mathcal{K}\|_2 \le \|\mathcal{K}\|_\infty$ in view of Assumption 4. Taking into account that $h_{\min} > n^{-1}$ we get

$$(4.26) \qquad R \le 8n\big(\Omega^{-1}(4r+1)\ln(n)\big)^{\frac{1}{\omega}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2\big[12Q + \sqrt{2}\big],$$

Note also that $\varkappa_1 \wedge \varkappa_2 \ge \underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2$ since $\Omega \le 1$. Thus, putting

$$c_2(n) = 730\ln\Big(16n^2(\Omega^{-1}(4r+1)\ln(n))^{\frac{1}{\omega}}\underline{g}^{-\frac{1}{2}}\big[12Q + \sqrt{2}\big]\Big) + 8r\ln(n) + 394,$$

we deduce from (4.25) and (4.26) that (4.12) holds with

$$\begin{aligned} U_n(\varkappa_1, \varkappa_2) \;&=\; 2\sqrt{2}(\sigma \vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_2(n)} \\ &\quad + (8/3)c_2(n)\big(\ln(n)\big)^{-\frac{1}{2}}\underline{g}^{-1}\|\mathcal{K}\|_\infty^2\big(\Omega^{-1}(4r+1)\big)^{\frac{1}{\omega}}c_2(n). \end{aligned}$$

It remains to note that $U_n(\varkappa_1, \varkappa_2) = C_2(n)$, our choice of τ provides $n\exp\{-\Omega\tau^\omega\} = n^{-r}$, and (4.23) follows from (4.24) with $T = C_2(n)$.

The assertion of the lemma follows now from (4.13) and (4.23). ∎

4.3. *Proof of Lemma 3.* We start the proof with the following statement. Set

$$\mathbb{F}_2(\beta_0, M) := \bigg\{W : \mathbb{R}^2 \to \mathbb{R} : \quad \|W\|_\infty + \sup_{\mathrm{y},\mathrm{y}' \in \mathbb{R}^2}\frac{|W(\mathrm{y}) - W(\mathrm{y}')|}{|\mathrm{y} - \mathrm{y}'|_2^{\beta_0}} \le M\bigg\}.$$

Then, for any $f \in \mathbb{F}(\beta_0, M)$ and any $\theta^* \in \mathbb{S}^1$ one has

$$(4.27) \qquad F \in \mathbb{F}_2(\beta_0, M),$$

for any F satisfying (1.2). Indeed, since the Cauchy-Schwartz inequality provides that $|(\mathrm{y} - \mathrm{y}')^\top\theta^*| \le |\mathrm{y} - \mathrm{y}'|_2$ for $\theta^* \in \mathbb{S}^1$, we have

$$\begin{aligned} &\|F\|_\infty + \sup_{\mathrm{y},\mathrm{y}' \in \mathbb{R}^2}\frac{|F(y_1) - F(y_2)|}{|\mathrm{y} - \mathrm{y}'|_2^{\beta_0}} = \|f\|_\infty + \sup_{\mathrm{y},\mathrm{y}' \in \mathbb{R}^2}\frac{|f(\mathrm{y}^\top\theta^*) - f(\mathrm{y}'^\top\theta^*)|}{|\mathrm{y} - \mathrm{y}'|_2^{\beta_0}} \\ &\le \|f\|_\infty + \sup_{y_1, y_2 \in \mathbb{R}}\frac{|f(y_1) - f(y_2)|}{|y_1 - y_2|^{\beta_0}}\sup_{\mathrm{y},\mathrm{y}' \in \mathbb{R}^2}\bigg(\frac{|\mathrm{y}^\top\theta^* - \mathrm{y}'^\top\theta^*|}{|\mathrm{y} - \mathrm{y}'|_2}\bigg)^{\beta_0} \le M, \end{aligned}$$

as a consequence of Assumption 3. The inclusion (4.27) implies

$$
\begin{aligned}
B_\infty &:= \sup_{t\in[-1/2,1/2]^2}\int_{\mathbb{R}^2}\Big|K_{\mathfrak{h}}(x-t)\big[F(x)-F(t)\big]\Big|\mathrm{d}x\\
&\le \|\mathcal{K}\|_1^2 M\mathfrak{h}^{\beta_0}\le\|\mathcal{K}\|_1^2, \qquad (4.28)
\end{aligned}
$$

in view of (1.6). Here we have also used that $\mathcal{K}$ is supported on $[-1/2,1/2]$.

Define $T(x,t)=\mathfrak{h}K_{\mathfrak{h}}(x-t)g^{-1}(x)$, and put for any $t\in[-1/2,1/2]^2$

$$
\begin{aligned}
\overline{\eta}_n(t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n\Big\{T(X_i,t)F(X_i)-\mathbb{E}_X^{(n)}\left[T(X_1,t)F(X_i)\right]\Big\},\\
\overline{\xi}_n(t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n T(X_i,t)\varepsilon_i.
\end{aligned}
$$

With this notation the standard "approximation + stochastic part" decomposition of the estimator $\widehat{F}(t)$ reads as follows:

$$
\widehat{F}(t) = \int_{\mathbb{R}^2}K_{\mathfrak{h}}(x-t)F(x)\mathrm{d}x+\big(n\mathfrak{h}^2\big)^{-\frac12}\left(\overline{\eta}_n(t)+\overline{\xi}_n(t)\right).
$$

That gives

$$
\big|\|\widehat{F}\|_\infty-\|F\|_\infty\big|\le\|\mathcal{K}\|_1^2+\big(n\mathfrak{h}^2\big)^{-\frac12}\left(\|\overline{\eta}_n\|_\infty+\|\overline{\xi}_n\|_\infty\right). \qquad (4.29)
$$

We will prove that

$$
\mathbb{P}_X^{(n)}\Big\{\|\overline{\eta}_n\|_\infty\ge C_3(n)\|F\|_\infty\Big\}\le 4n^{-4r}, \qquad (4.30)
$$

$$
\mathbb{P}_{X,\varepsilon}^{(n)}\Big\{\|\overline{\xi}_n\|_\infty\ge C_4(n)\Big\}\le(4+\Upsilon)n^{-4r}, \qquad (4.31)
$$

where $C_3(n)$ an $C_4(n)$ are given in Section 3.1.

Let us show how to deduce the assertion of the lemma from (4.30) and (4.31). Indeed, remembering that $C_5(n)=\|\mathcal{K}\|_1^2+\big(n\mathfrak{h}^2\big)^{-\frac12}C_4(n)$ and that $C_3(n)\big(n\mathfrak{h}^2\big)^{-\frac12}\le 1/2$ for any $n\ge n_1$ in view of restriction (2.6), we obtain from (4.29)

$$
\begin{aligned}
&\Big\{\|\overline{\eta}_n\|_\infty<C_3(n)\|F\|_\infty\Big\}\cap\Big\{\|\overline{\xi}_n\|_\infty<C_4(n)\Big\}\\
\subseteq\ &\Big\{\big|\|\widehat{F}\|_\infty-\|F\|_\infty\big|\le C_5(n)+2^{-1}\|F\|_\infty\Big\}\\
=\ &\Big\{\widehat{F}_\infty\in\Big[\|F\|_\infty, 3M+4C_5(n)\Big]\Big\}.
\end{aligned}
$$

It yields for any $F \in \mathbb{F}_2(\beta_0, M)$

$$
\begin{aligned}
& \mathbb{P}_F^{(n)}\Big\{ \widehat{F}_\infty \notin \Big[\|F\|_\infty, 3M + 4C_5(n)\Big] \Big\} \\
\leq \;& \mathbb{P}_X^{(n)}\Big\{ \|\overline{\eta}_n\|_\infty \geq C_3(n)\|F\|_\infty \Big\} + \mathbb{P}_{X,\varepsilon}^{(n)}\Big\{ \|\overline{\xi}_n\|_\infty \geq C_4(n) \Big\},
\end{aligned}
$$

and the assertion of the lemma follows from (4.30) and (4.31) since the right hand side of the inequality is independent of θ^* and f.

Thus, let us proceed with the proof of (4.30). It is based on the application of inequality (4.12) established in the proof of Lemma 2.

1^0. We apply (4.12) with $\zeta_n(z) = \overline{\eta}_n(t))$, $z = t$, $\mathbb{V} = \mathbb{R}^2$, $\mathfrak{s} = 2$, $B = 1$, $\mathcal{Z} = [-1/2, 1/2]^2$ and $G(\cdot, z) = T(\cdot, t)F(\cdot)$.

Since K is supported on $[-1/2, 1/2]^2$ and $\mathfrak{h} \leq 1$ for any $n \geq n_0$ we obtain

$$
T(x,t) = 0, \quad \forall x \neq [-1,1]^2, \;\; \forall t \in [-1/2, 1/2]^2. \tag{4.32}
$$

It yields together with Assumption 1 and (4.27) that

$$
G_2^* \leq \|F\|_\infty \underline{g}^{-\frac{1}{2}} \|\mathcal{K}\|_\infty^2 =: \varkappa_1, \qquad G_\infty^* \leq \mathfrak{h}^{-1}\|F\|_\infty \underline{g}^{-1}\|\mathcal{K}\|_\infty^2 =: \varkappa_2. \tag{4.33}
$$

We see that (4.6) is fulfilled. Now let us check (4.7). We have in view of (4.32) and the triangle inequality for any $t, t' \in [-1/2, 1/2]^2$

$$
\begin{aligned}
b_\infty(t, t') \;&:=\; \sup_{x \in [-1,1]^2} \big| T(x,t) - T(x,t') \big| \, |F(x)| \\
&\leq\; \mathfrak{h}\|F\|_\infty \underline{g}^{-1} \sup_{x \in [-1,1]^2} \big| K_\mathfrak{h}(x-t) - K_\mathfrak{h}(x-t') \big| \\
&\leq\; 2\mathfrak{h}^{-2}\|F\|_\infty Q \underline{g}^{-1}\|\mathcal{K}\|_\infty \big| t - t' \big|_\infty.
\end{aligned} \tag{4.34}
$$

The latter inequality follows from Assumption 4. Thus, we obtain that (4.7) holds with

$$
\alpha = 1, \qquad R = 2n^2 Q \underline{g}^{-1}\|\mathcal{K}\|_\infty^2 \|F\|_\infty. \tag{4.35}
$$

Here we have used that $\mathfrak{h} \geq n^{-1}$ and $\|\mathcal{K}\|_\infty \geq 1$. Thus, putting

$$
c_3(n) = 365 \ln\Big(2n^2 Q \underline{g}^{-\frac{1}{2}}\Big) + 8r\ln(n) + 197,
$$

we deduce from (4.33) and (4.35) that (4.12) holds with

$$
U_n(\varkappa_1, \varkappa_2) \;=\; C_3(n)\|F\|_\infty,
$$

where remind, $C_3(n)=2\sqrt{2}\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|_\infty^2\sqrt{c_3(n)}+(8/3)\underline{g}^{-1}\|\mathcal{K}\|_\infty^2 c_3(n)\big(n\mathfrak{h}^2\big)^{-\frac{1}{2}}$. Thus, (4.30) is established.

2^0. To prove the inequality (4.31) we first note that similarly to (4.24) we have for any $y>0$ and any $\tau>0$

$$\begin{aligned}&\mathbb{P}^{(n)}_{X,\varepsilon}\bigg(\sup_{t\in[-1/2,1/2]^2}\big|\overline{\xi}_n(t)\big|\geq y\bigg)\\ &\leq \mathbb{P}^{(n)}_{X,\varepsilon}\bigg(\sup_{t\in[-1/2,1/2]^2}\big|\overline{\xi}^{(1)}_n(t)\big|\geq y\bigg)+n\Upsilon e^{-\Omega\tau^\omega}\end{aligned}\tag{4.36}$$

in view of Assumption 1. Here we have denoted

$$\overline{\xi}^{(1)}_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^n T(X_i,t)\varepsilon_i 1_{[-\tau,\tau]}(\varepsilon_i).$$

We will apply (4.12) with $\zeta_n(z)=\overline{\xi}^{(1)}_n(t),\ z=t,\ \mathcal{Z}=[-1/2,1/2]^2$, $\mathfrak{s}=2$, $B=1$ and $G(v,z)=T(x,t)w1_{[-\tau,\tau]}(w),\ v=(x,w)\in\mathbb{V}:=\mathbb{R}^3$.

We have in view of Assumption 1 and (4.32)

$$G^*_2\leq(\sigma\vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|^2_\infty=:\varkappa_1,\qquad G^*_\infty\leq\tau\mathfrak{h}^{-1}\underline{g}^{-1}\|\mathcal{K}\|^2_\infty=:\varkappa_2,\tag{4.37}$$

and, therefore, (4.6) holds.

Using the last inequality in (4.34) we obtain for any $t,t'\in[-1/2,1/2]^2$

$$\begin{aligned}b_\infty\big(t,t'\big)&:=\sup_{(x,w)\in\mathbb{R}^2}\big|T(x,t)-T(x,t')\big|\,|w|1_{[-\tau,\tau]}(w)\\ &\leq 2\tau\mathfrak{h}^{-2}Q\underline{g}^{-1}\|\mathcal{K}\|_\infty\big|t-t'\big|_\infty.\end{aligned}$$

Hence, (4.7) is fulfilled with

$$\alpha=1,\qquad R=2\tau n^2Q\underline{g}^{-1}\|\mathcal{K}\|^2_\infty.\tag{4.38}$$

Choose $\tau=\big(\Omega^{-1}(4r+1)\ln(n)\big)^{\frac{1}{\omega}}$ and note that $\varkappa_1\wedge\varkappa_2\geq\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|^2_\infty$ since $\Omega\leq 1$. Thus, putting

$$c_4(n)=365\ln\Big(2n^2(\Omega^{-1}(4r+1)\ln(n))^{\frac{1}{\omega}}\underline{g}^{-\frac{1}{2}}Q\Big)+8r\ln(n)+197,$$

we deduce from (4.37) and (4.38) that (4.12) holds with

$$U_n(\varkappa_1,\varkappa_2)=2\sqrt{2}(\sigma\vee 1)\underline{g}^{-\frac{1}{2}}\|\mathcal{K}\|^2_\infty\sqrt{c_4(n)}+(8/3)\tau c_4(n)\big(n\mathfrak{h}^2\big)^{-\frac{1}{2}}\underline{g}^{-1}\|\mathcal{K}\|^2_\infty.$$

Since $U_n(\varkappa_1,\varkappa_2)=C_4(n)$, choosing in (4.36) $y=C_4(n)$, we come to (4.31).

■

4.4. *Proof of Lemma 5.* First, in view of the $(\mathfrak{p},\mathfrak{p})$-strong maximal inequality, see e.g. Theorem 9.16 in Wheeden and Zygmund (1977), one has

$$\left\|\Delta^{*}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}} \le \tau_{\mathfrak{p}} \left\|\Delta_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}},$$

where the constant $\tau_{\mathfrak{p}}$ depends only of $\mathfrak{p}$.

For any $\delta\in(0,h]$ put $B(z,\delta)=\left|\delta^{-1}\int\mathcal{K}\big([u-z]/\delta\big)\big(g(u)-g(z)\big)\mathrm{d}u\right|$ and define

$$\Delta^{(n)}_{\mathcal{K},g}(h,z)=\sup_{\delta\in[hn^{-1},h]}B(z,\delta),\quad n=1,2,\dots\,.$$

We remark that the sequence $\{\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\}_{n\ge1}$ increases monotonically and $\Delta^{(n)}_{\mathcal{K},g}(h,z)\to\Delta_{\mathcal{K},g}(h,z)$ for any $z\in\mathbb{R}$, as $n\to\infty$. Hence, by Beppo-Levi's theorem

$$\left\|\Delta_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}}=\lim_{n\to\infty}\left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}},$$

and, in view of (4.4), to complete the argument we need to show that

$$\sup_{g\in\mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})}\left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}}\le 2\mathcal{Q}h^{s}\|\mathcal{K}\|_{\infty}\left[2^{s\mathfrak{p}}-1\right]^{-\frac{1}{\mathfrak{p}}},\quad\forall n\ge1. \tag{4.39}$$

Assumption 4 (*2*) implies that we can assert that $B(z,\cdot)$ is continuous on $[n^{-1}h,h]$. Hence for any $z\in\mathbb{R}$ there exists $\delta(z)\in[n^{-1}h,h]$ such that

$$\Delta^{(n)}_{\mathcal{K},g}(h,z)=B\big(z,\delta(z)\big). \tag{4.40}$$

For any $l=0,\dots,\log_2 n-1$ (without loss of generality $\log_2 n$ is assumed an integer) we consider the slices $V_l=\big\{z\in\mathbb{R}:\ \ a_{l+1}<\delta(z)\le a_l\big\}$ with $a_l=2^{-l}h$. Later on the integration over empty set is supposed to be zero. Then

$$\left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}}^{\mathfrak{p}}=\sum_{l=0}^{\log_2 n-1}\int_{V_l}\left|B\big(z,\delta(z)\big)\right|^{\mathfrak{p}}\mathrm{d}z. \tag{4.41}$$

We will treat the cases $s\le1$ and $s>1$ separately. If $s<1$, on any slice V_l, $l=0,\dots,\log_2 n$, we have

$$\begin{aligned}
B\big(z,\delta(z)\big) &\le \frac{\|\mathcal{K}\|_{\infty}}{\delta(z)}\int_{-\frac{\delta(z)}{2}}^{\frac{\delta(z)}{2}}|g(z+v)-g(z)|\,\mathrm{d}v\\
&\le \frac{2\|\mathcal{K}\|_{\infty}}{a_l}\int_{-\frac{a_l}{2}}^{\frac{a_l}{2}}|g(z+v)-g(z)|\,\mathrm{d}v\\
&= 2\|\mathcal{K}\|_{\infty}\int_{-\frac12}^{\frac12}|g(z+ta_l)-g(z)|\,\mathrm{d}t. \qquad (4.42)
\end{aligned}$$

We obtain from (4.41) and (4.42) with the use of Minkowski's inequality for integrals and writing for ease of notation $\mu = 2\|\mathcal{K}\|_\infty$ that

$$\begin{aligned}
\left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|^{\mathfrak{p}}_{\mathfrak{p}} &\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} \int \left| \int_{-\frac{1}{2}}^{\frac{1}{2}} |g(ta_l+z)-g(z)|\,\mathrm{d}t \right|^{\mathfrak{p}} \mathrm{d}z \\
&\leq \mu^{\mathfrak{p}} \sum_{l=0}^{\log_2 n-1} \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \|g(\cdot+ta_l)-g(\cdot)\|_{\mathfrak{p}}\,\mathrm{d}t \right)^{\mathfrak{p}} \\
&\leq \left[\frac{\mathcal{Q}h^s\|\mathcal{K}\|_\infty 2^{1-s}}{(s+1)} \right]^{\mathfrak{p}} \sum_{l=0}^{\infty} 2^{-ls\mathfrak{p}}.
\end{aligned}$$

Here we have used that $g \in \mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})$. Thus, we have for any $s \leq 1$ and any $n \geq 1$

$$\sup_{g\in\mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})} \left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}} \leq 2\mathcal{Q}h^s\|\mathcal{K}\|_\infty \left[2^{s\mathfrak{p}}-1\right]^{-\frac{1}{\mathfrak{p}}}. \tag{4.43}$$

If $s > 1$, using Taylor's formula we have for any $g \in \mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})$ any $v \in \mathbb{R}$

$$\begin{aligned}
g(v+z)-g(z) &= \sum_{m=1}^{m_s} \frac{g^{(m)}(z)}{m!} v^m \\
&\quad + \frac{v^{m_s}}{(m_s-1)!} \int_0^1 (1-\lambda)^{m_s-1} \left[g^{(m_s)}(z+v\lambda) - g^{(m_s)}(z) \right] \mathrm{d}\lambda.
\end{aligned}$$

We have in view of Assumptions 4 and 5 for any $z \in \mathbb{R}$

$$\begin{aligned}
B\big(z,\delta(z)\big) &\leq \frac{\|\mathcal{K}\|_\infty}{(m_s-1)!} \frac{1}{\delta(z)} \\
&\quad \int_{-\frac{\delta(z)}{2}}^{\frac{\delta(z)}{2}} \int_0^1 |v|^{m_s} (1-\lambda)^{m_s-1} \left| g^{(m_s)}\big(z+\lambda v\big) - g^{(m_s)}(z) \right| \mathrm{d}\lambda \mathrm{d}v.
\end{aligned}$$

By the latter inequality for any $z \in V_l$ we get

$$\begin{aligned}
& B\big(z,\delta(z)\big) \\
&\leq \frac{2\|\mathcal{K}\|_\infty a_l^{m_s}}{(m_s-1)!} \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^1 |t|^{m_s} (1-\lambda)^{m_s-1} \left| g^{(m_s)}(z+\lambda t a_l) - g^{(m_s)}(z) \right| \mathrm{d}\lambda \mathrm{d}t.
\end{aligned} \tag{4.44}$$

Thus, we obtain from (4.40), (4.41) and (4.44) with the use of Minkowski's

inequality for integrals and denoting $\mu = 2\|\mathcal{K}\|_\infty/(m_s-1)!$ that

$$
\begin{aligned}
\left\|\Delta^{(n)}_{\mathcal{K},f}(h,\cdot)\right\|^{\mathfrak{p}}_{\mathfrak{p}} &= \sum_{l=0}^{\log_2 n-1}\int_{V_l}|B\big(z,\delta(z)\big)|^{\mathfrak{p}}\mathrm{d}z \\
&\le \mu^{\mathfrak{p}}\sum_{l=0}^{\log_2 n-1} a_l^{m_s\mathfrak{p}} \\
&\quad\times\int\left(\int_{-\frac12}^{\frac12}\int_0^1 |t|^{m_s}(1-\lambda)^{m_s-1}\left|g^{(m_s)}(z+\lambda t a_l)-g^{(m_s)}(z)\right|\mathrm{d}\lambda\mathrm{d}t\right)^{\mathfrak{p}}\mathrm{d}z \\
&\le \mu^{\mathfrak{p}}\sum_{l=0}^{\log_2 n-1} a_l^{m_s\mathfrak{p}} \\
&\quad\times\left(\int_{-\frac12}^{\frac12}\int_0^1 |t|^{m_s}(1-\lambda)^{m_s-1}\left\|g^{(m_s)}(\cdot+\lambda t a_l)-g^{(m_s)}(\cdot)\right\|_{\mathfrak{p}}\mathrm{d}\lambda\mathrm{d}t\right)^{\mathfrak{p}} \\
&\le \left[\frac{\mathcal{Q}h^s\|\mathcal{K}\|_\infty 2^{1-s}}{(s+1)(m_s+1)(m_s-1)!}\right]^{\mathfrak{p}}\sum_{l=0}^{\infty}2^{-ls\mathfrak{p}}.
\end{aligned}
$$

Here we have used that $g\in\mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})$. Thus, we have for any $s>1$ and $n\ge1$

$$
\sup_{g\in\mathbb{N}_{\mathfrak{p}}(s,\mathcal{Q})}\left\|\Delta^{(n)}_{\mathcal{K},g}(h,\cdot)\right\|_{\mathfrak{p}} \le 2\mathcal{Q}h^s\|\mathcal{K}\|_\infty\left[2^{s\mathfrak{p}}-1\right]^{-\frac{1}{\mathfrak{p}}}. \tag{4.45}
$$

We conclude that (4.39) is established in view (4.43) and (4.45). ∎

References.

Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM P &S* 6, 127–146. MR1918295.

Besov, O. V., Il'in, V. P. and Nikol'skii, S. M. (1978, 1979). *Integral Representations of Functions and Imbedding Theorems.*, Vol. I,II. Scripta Series in Mathematics., V. H. Winston & Sons, Washington, D.C.; Halsted Press [John Wiley & Sons], New York-Toronto, Ont.-London. MR0519341, MR0521808.

Brown, L.D. and Low, M.G. (1996). A constrained risk inequality with application to nonparametric functional estimation. *Annals Statist.* **24**, 6, 2524–2535. MR1425965.

Chesneau, C. (2007). Regression with random design: minimax study. *Statist. Probab. Lett* **77**, 1, 40–53. MR2339017.

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.* **3:3**, 215–228. MR1400080.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wawelet shrinkage: asymptopia?*J. Roy. Statist. Soc.* Ser. R **57** 301–369. MR1323344.

Jennrich, R. (1969). Asymptotic properties of non–linear least squares estimators. *Ann. Math. Statist.* **40**, 633–643. MR0238419.

Gaïffas, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM P &S* 11, 334–364. MR2339297.

Gaïffas S. and Lecué G. (2007). Optimal rates and adaptation in the single-index model using aggregation. *Electronic J. Statist.* 1, 538–573. MR2369025.

Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14:4** 1150–1190. MR2543590.

Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via L_p-norm oracle inequalities. *Probab. Theory Related Fields* **143:1-2** 41–71. MR2449122.

Golubev, G. K. (1992). Asymptotically minimax estimation of a regression function in an additive model. *Problems Inform. Transmission* **28:2** 101–112. MR1178413.

Horowitz J. L. (1998). *Semiparametric and Nonparametric Methods in Econometrics.* Lecture Notes in Statistics, 131. Springer-Verlag, New York, 1998. MR1624936.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Benoulli* **10**, 6, 1053–1105. MR2108043.

Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Nonlinear estimation in anisotropic multi–index denoising. *Probab. Theory Related Fields* **121**, 137–170. MR1863916.

Kerkyacharian, G., Lepski, O. and Picard, D. (2008). Nonlinear estimation in anisotropic multi–index denoising. Sparce case. *Probab. Theory Appl.* **52**, 150–171. MR2354574.

Kulik R. and Raimondo M. (2009). Wavelet regression in random design with heteroscedastic dependent errors. *Annals Statist.* **37**, 6, 3396–3430. MR2549564.

Lepskii, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35:3** 454–466. MR1091202.

Lepski, O. (2013). Upper Functions for Positive Random Functionals. I. General Setting and Gaussian Random Functions, *Math. Methods Statist.* **22:1**, 1–27.

Lepski, O. V., Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25:3** 929–947. MR1447734.

Maddala, G. (1983). *Limited-Dependent and Qualitative Variables in Econometrics.* Econometric Society Monographs in Quantitative Economics, 3. Cambridge University Press, Cambridge, 1983. MR0799154.

Zhang, S., Wong, M-Yu. and Zheng, Z. (2002). Wavelet threshold estimation of a regression function with random design. *J. Multiv. Anal.* **80**, 2, 256–284. MR1889776.

Wheeden, R. L. and Zygmund, A. (1977). *Measure and integral. An introduction to real analysis.* Pure and Applied Mathematics, Vol. 43. Marcel Dekker, Inc., New York-Basel. MR0492146.

Laboratoire d'Analyse,
Topologie, Probabilités UMR 7353
Aix-Marseille Université
39, rue F. Joliot Curie
13453 Marseille FRANCE
E-mail: Oleg.Lepski@cmi.univ-mrs.fr

Institute for Mathematical Stochastics
Georg-August-Universität Göttingen
Goldschmidtstrasse 7
37077 Göttingen GERMANY
E-mail: Nora.Serdyukova@gmail.com